Learning to abstract

Ley, A.

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

LEARNING TO ABSTRACT

How short-term experience shapes neural sound representations

Anke Ley

LEARNING TO ABSTRACT

How short-term experience shapes
neural sound representations

Anke Ley

Cover Anke Ley
Production CPI Wöhrmann Print Services B.V.
ISBN 978-94-6203-455-6

LEARNING TO ABSTRACT

How short-term experience shapes neural sound representations

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan Tilburg University
op gezag van de rector magnificus,
prof. dr. Ph. Eijlander,
in het openbaar te verdedigen ten overstaan van een
door het college voor promoties aangewezen commissie
in de aula van de Universiteit

op woensdag 23 oktober 2013 om 14.15 uur

door

Anke Ley

geboren op 9 december 1983 te Los Alamos,
Verenigde Staten van Amerika

PROMOTIECOMMISSIE

Promotores

Prof. dr. J.H.M. Vroomen
Prof. dr. E. Formisano

Overige leden

Prof. dr. J.J. Foxe
Prof. dr. A.J. van Opstal
Dr. N. van Atteveldt
Dr. F. De Martino
Dr. L. Riecke

Chapter 1

INTRODUCTION

Wisdom is learning what to overlook.

~ William James

Auditory neuroscience has made major advances in the past years to unravel the functional organization principles of the auditory system. Most research in this domain has focused on the mechanisms underlying the neural representation[1] of acoustic sound properties. The ability to precisely process and uniquely represent each and every sound is however insufficient to accurately interact with the tremendous diversity of the natural auditory environment. To process sounds in an efficient and goal-directed manner, it is important to abstract from the explicit representation of the acoustic signal by ignoring irrelevant variability. Crucially, what is relevant or irrelevant is determined by the current situation and goal and might therefore change over time. The same sound may be interpreted in different ways according to the context in which it is perceived and the environmental requirements. This places a high demand on the neural representation of sounds, as they need to bear the flexibility to rapid strategic modifications rather than being rigid mappings of the physical sound properties. Abstract and behaviorally relevant sound representations are the basis for cognitive abilities such as recognition and categorization, yet their neural implementation remains largely unknown.

The experimental work presented in this thesis investigates the transformation from the acoustic representation of sounds into a meaningful context-dependent perceptual representation. To this end, psychophysics and functional magnetic resonance imaging (fMRI, see below for an explanation) techniques are employed. The aim of this research is to identify those brain areas mediating the abstraction of behaviorally relevant representations from the exact mapping of the acoustic sound features as a function of short-term experience. For this purpose, all studies presented in the following relied on the comparison of sound representations prior to and following successful learning of novel sound categories. Categories are natural and meaningful groupings of acoustically variant sounds and

1 Neural representation: how the structured spatiotemporal pattern of neuronal activity reflects the characteristics of the stimulus or of the percept (Eggermont, 2001)

therefore present optimal candidates to investigate perceptual abstraction processes. As the studies in this thesis aim to elucidate the role of the dynamic interaction between the acoustic input and environmental factors, such as context, feedback, and prior experience in shaping the internal sound representations, they employ artificial non-speech sounds without attached meaning or behavioral relevance prior to learning. Furthermore, this research provides a unique comparison of audio-tactile and audio-visual learning conditions to investigate how naturally occurring multisensory experience shapes sound representations.

This chapter begins by providing essential background information on the concept of categorization and illustrates the phenomenon of Categorical Perception (Harnad, 1987). The remainder of this chapter will review existing research on the neural basis of natural categories with a special emphasis on the function of learning and plasticity in the translation of physical sensory information into functionally relevant representations. Finally, this chapter will discuss the prevalence of crossmodal plasticity and provide an outline of the subsequent chapters.

1 | SOUND PROCESSING – MORE THAN TIME-FREQUENCY ANALYSIS

Sounds are pressure waves travelling through the air. Each and every sound wave is characterized by unique physical characteristics determined by the frequency of the pressure oscillations, their amplitude and the dynamic changes of these two features over time. The human auditory system is perfectly equipped to process these non-stationary acoustic signals, starting from a complete frequency analysis at the level of the cochlea[2] in the inner ear (Kandel et al., 2000) to more complex spectro-temporal processing properties of neuronal populations in the auditory cortex (Kowalski et al., 1996a, 1996b; Schönwiesner & Zatorre, 2009). Yet, these complete representations of the acoustic sound parameters alone cannot bear basic perceptual abilities such as sound recognition and categorization. Unless the context in which a sound is repeated is absolutely identical to the first encounter - which is rather unlikely under natural circumstances - recognizing a sound as 'the same' is not trivial, given that the exact acoustic profiles of the two repetitions may not entirely match. Examples of sources of variability are differences in the anatomy of the vocal tract of different speakers pronouncing the same word, naturally occurring background noise and obstacles obstructing the path from sound source to receiver potentially muting or altering the original waveform. Even the same word uttered by the same speaker twice would create non-identical spectral profiles (Klein & Zatorre, 2011) and hence result in a mismatch between the input and a stored template. How then to distinguish the repetition of the same sound from two different sounds?

Object recognition requires perceptual constancy

Categorization presents an even bigger challenge to the sound representation, as categories may consist of acoustically different sounds

[2] Cochlea: spiral-shaped bony structure in the inner ear containing the basilar membrane with different resonant frequencies along its length that transmits the sound-driven vibrations to the sensory hair cells

that should nevertheless be treated alike. Just as not all chairs look identical, the sounds of birds differ to a large degree (take for example a crow, a duck, a pigeon, a seagull, and a mockingbird, see **Figure 1.1** for the corresponding time-frequency spectrograms) but humans learned to distinguish the large variety of bird sounds from the sounds of other animals without effort. Conversely, humans manage to separate acoustically very similar sounds such as the bleating of two different goats and the human imitation of the same sound. As these examples show, the inability to abstract from the uniqueness of each individual stimulus would pose extreme challenges to ordinary life and reflects a marked characteristic of pathological disorders such as the autism spectrum (Minshew et al., 2002). Furthermore, reducing the mental representation of an acoustic signal to its relevant information content does not only establish the basis for meaningful interaction with the outside world, it also essentially decreases the need for processing resources (Olshausen & Field, 2004).

However, despite the ease with which humans generally accomplish this task, the detection of relevant and invariant information in the complexity of the sensory input is not trivial. In the visual domain, for instance, non-uniform color and illumination conditions in natural environments cause major difficulties for sensor-equipped robots in object recognition tasks and no simple algorithms manage to solve this 'invariance problem' created by real-life situations (Sridharan & Stone, 2009). Along the same lines, the performance of voice recognition systems for human-computer interaction is far below that of humans, suffering mainly from the naturally occurring variability in speech signals (Benzeguiba et al., 2007).

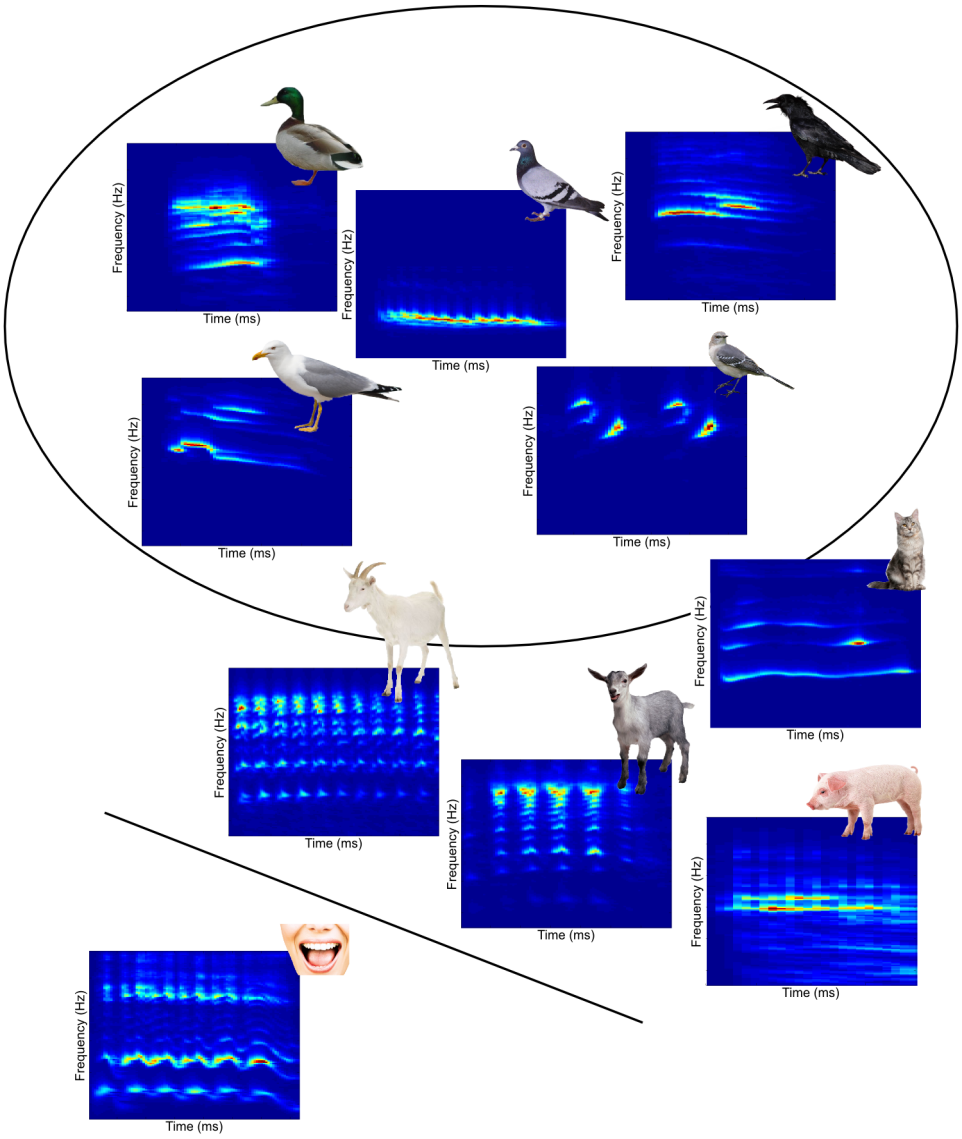


Figure 1.1 Time-frequency spectrograms for a sample of natural sound categories. Despite large acoustic variability among bird sounds, humans categorize them as such without effort. Acoustically similar sounds as for example the bleating of goats and the imitation of the same sound produced by a human speaker are easily distinguished. Color temperature indicates power in the respective time-frequency bin.

2 | PERCEPTUAL CATEGORIES AND 'CATEGORICAL PERCEPTION'

Categories are groupings of objects that are considered equivalent (Rosch, 1998). The exact definition of categories is not trivial, given the different abstraction levels in the taxonomy. Three different levels of abstraction are commonly specified, namely superordinate categories (e.g. animals, vehicles), basic categories (e.g. dog, cat), and subordinate categories (e.g. collie, mastiff). While superordinate categories can be easily discriminated based on their global physical attributes (e.g. 'has wheels'), the overlap of features increases when decreasing the abstraction level. Hence, subordinate categories can only be discriminated based on few discriminative features, while sharing a multitude of other properties.

It is quite remarkable that despite the ease with which even subordinate categories are generally discriminated, there are often no natural discontinuities in the underlying physical structure of objects dictating where the category boundary lies. Instead, categories are perceived as distinct despite continuous variation in their underlying physical attributes. Colors for instance, are perceived as discrete qualities such as red, blue, green, or yellow. The underlying visual property of the stimulus (i.e. the light spectrum) however, varies in a continuous fashion with no physical indication for separate color ranges (Bornstein, 1987). Perceptually, the difference between green and yellow is identified instantly, while two wavelengths differing by the same amount within the range of what is called green look like varying shades of the same color (see **Figure 1.2A**).

In the auditory domain, speech categories are characterized by the same discrepancy between physical and perceptual similarity: The perceptual boundaries between phoneme categories such as /b/, /d/, and /g/ **(Figure 1.2B**, here displayed as two-syllable utterances /ba/, /da/, and /ga/) are completely arbitrary in view of the fact that the underlying acoustic features vary smoothly from one category to the next, giving no hint where one phoneme starts and the other ends. Remarkably though, if people are asked to identify individual sounds randomly taken from this phoneme continuum as either /ba/, /da/, or /ga/, their percept does not vary gradually as suggested

by the sensory input. Instead, the first portion of the continuum is robustly identified as /ba/, the middle portion as /da/, and the remaining part as /ga/ with abrupt perceptual switches in between (**Figure 1.2C**). This indicates that people are unaware of the underlying variation of the stimuli within one phoneme category, mapping various physically different stimuli onto the same identity. At the category boundary, however, the same extent of physical difference is perceived as a change in stimulus identity. This difference in discrimination ability suggests that the within-category differences in the physical domain are perceptually compressed to create a robust representation of the phoneme or color category while between-category differences are perceptually enhanced to detect the relevant change of phoneme or color identity. This phenomenon is termed **'**Categorical Perception' (CP, Harnad, 1987) and has been demonstrated for stimuli from various natural domains apart from speech and color, such as music (Burns & Ward, 1978) and facial expressions of emotion (Etcoff & Magee, 1992) and not only for humans but monkeys (Freedman et al., 2001, 2003), chinchillas (Kuhl & Miller, 1975), songbirds (Prather et al., 2009) and even crickets (Wyttenbach et al., 1996). Thus, it seems to be a universal reduction mechanism to deal with the complexity of natural environments.

It can be assumed that apart from few exceptions, most natural categories are learned through experience

~ Stevan Harnad

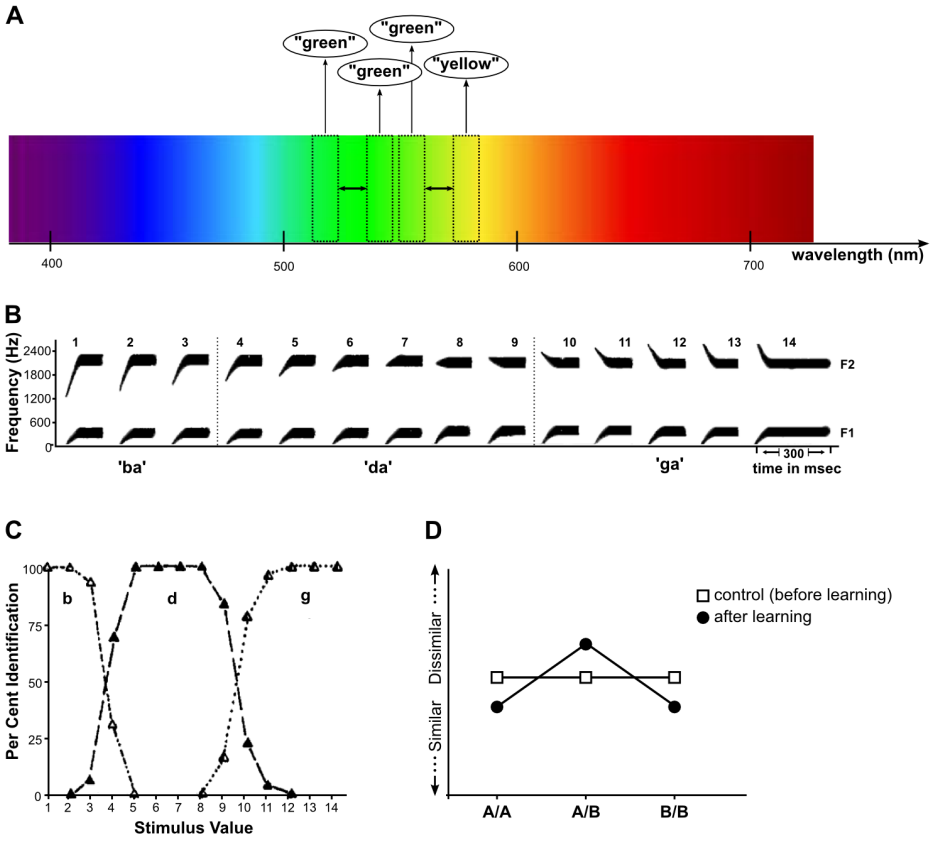


Figure 1.2 Illustration of the perceptual transformation in categorical perception. A, Color spectrum of the visible range. Squares depict equally spaced wavelength distances that are easily discriminated between color categories (green/yellow) and almost indistinguishable within a category (green). **B**, Schematic representation of spectral patterns for the continuum between the phonemes /b/, /d/, and /g/. F1 and F2 reflect the first and second formant (i.e. amplitude peaks in the frequency spectrum). **C**, Phoneme identification curves corresponding to the continuum in B. Curves are characterized by relatively stable percepts within a phoneme category and sharp transitions in between. B and C adapted from Liberman et al. (1957) **D**, Schematic illustration of learning-induced similarity ratings for exemplars within categories (A/A and B/B) and across category boundaries (A/B), reflecting 'acquired equivalence' and 'acquired distinctiveness' (Goldstone, 1994). Graphic adapted from Livingston et al. (1998).

3 | NATURE OR NURTURE?

The question that naturally arises from the description of CP is: *How* are perceptual categories formed? According to the view of psychological nativism, whose most famous representative is philosopher and cognitive scientist Jerry Fodor, all categories are innate, i.e. no learning is required to shape the mental representations of the sensory input such that we perceive it categorically but we have the intrinsic capacity to detect the invariances of the outside world (Fodor, 1975). It may be argued that color is an example of these innate perceptual categories because it is the properties of sensory receptors and retinal cells in the visual system that selectively respond to certain ranges of wavelength and automatically partition the input into distinct color categories. The view that color categories are universal is supported by findings of categorical color perception in pre-linguistic infants (Bornstein et al., 1976), and the consistency of color categories across adults with different cultural backgrounds and native languages (Bornstein, 1987). However, the influence of learning and language (i.e. linguistic labels for the color ranges) in forming perceptual color categories is not fully ruled out and still a matter of current investigation (Roberson et al., 2000; Gilbert et al., 2006; Brown et al., 2011).

Similarly to color categories, the relative contribution of innate processing mechanisms and learning in the formation of speech categories is not completely resolved. Despite the striking consistency of perceptual phoneme boundaries across different listeners, those boundaries are malleable depending on the context in which the sounds are perceived (Benders et al., 2010). Additionally, cross-cultural studies have shown that language learning influences the discriminability of speech sounds, such that phonemes in one particular language are only perceived categorically by native speakers of that language and continuous otherwise (Goldstone, 1994). In support of the claim that speech CP can be acquired through training stand experimental learning studies that successfully induced

discontinuous perception of a non-native phoneme continuum through elaborate category training (Myers & Swan, 2012).

To test for perceptual changes induced by category learning, the within- and between-category contrast sensitivity is typically assessed with perceptual discrimination tasks. Strictly, CP can only be assumed if the discrimination ability of items at the category boundary is enhanced relative to the discrimination of items that fall within the same category (Goldstone, 1994). Notably, this can be achieved in two (non-exclusive) ways (see **Figure 1.2D**): First, sensitivity for between-category contrasts can be increased with learning, second, sensitivity for within-category contrasts can be decreased with learning also called 'perceptual magnet effect' (Kuhl, 1991). These processes, referred to as 'acquired distinctiveness' and 'acquired similarity' respectively are difficult to tease apart and categorical effects have been demonstrated in combination with one as well as both sensitivity effects present, suggesting that they are mediated by different mental operations. At this point, it is important to note that the experimental studies conducted in the context of this thesis employ a less strict behavioral criterion for learning-induced categorical perception effects based on the sharp boundaries of the stimulus identification curves, similar to the characteristic effects observed for phoneme categories (Liberman et al., 1957, see **Figure 1.2C**).

A prevalent theory for the formation of speech categories in particular is the motor theory of speech perception (Liberman & Mattingly, 1985). This theory claims that speech sounds are categorized based on the distinct motor commands for the vocal tract used for pronunciation. Further fueled by the discovery of mirror neurons[3], the theory still has its proponents (for review see Galantucci et al., 2006), however, today, it is rejected in its strict form in which speech processing is considered special, as the recruitment of the motor system for sound identification has been demonstrated for various

[3] Mirror neurons: neurons that respond to both the action as well as the observation of the same action performed by someone else (for review, see Rizzolatti & Craighero (2004).

forms of non-speech action-related sounds (Kohler et al., 2002). Furthermore, accumulating evidence indicates that categorical perception can be induced by learning for a variety of non-speech stimulus material. In the visual domain, for instance, Goldstone (1994) presented multiple highly controlled experimental studies, in which he revealed changes in the discriminability of simple shapes through category training. Using more complex artificial objects, Livingston and colleagues (Livingston et al., 1998) also demonstrated effects of categorical perception induced by learning. Along similar lines, learning-induced characteristics of categorical perception have been reported after categorization training for simple noise sounds (Guenther et al., 1999) and inharmonic tone complexes (Goudbeek et al., 2009). The use of artificially constructed categories has the advantage that the physical distance between neighboring stimuli can be controlled such that the similarity ratings of within- or between-category stimuli can be attributed to true perceptual effects, rather than the metrics of the stimulus dimensions (see **Figure 1.3**).

We don't put items in the same category because they look alike - they look alike because we put them in the same category!

~ adapted from Livingston et al., (1998)

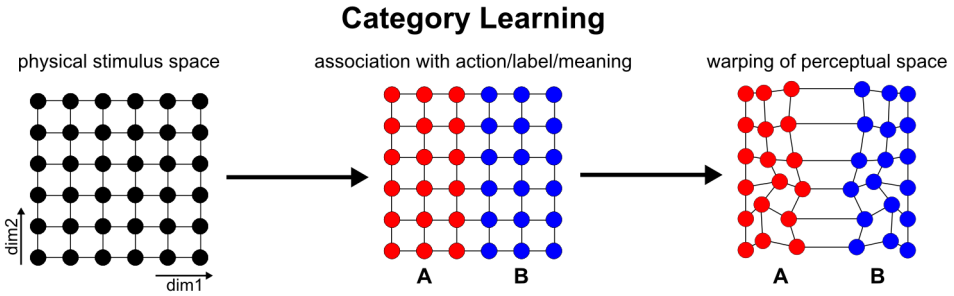


Figure 1.3 Schematic representation of the perceptual (and possibly neural) transformation from a continuum to a discrete categorical representation. The first plot depicts an artificial two-dimensional stimulus space without physical indications of a category boundary (exemplars are equally spaced along both dimensions). During learning, stimuli are separated according to the relevant dimension, irrespective of the variability in the second dimension. Lasting differential responses for the left and right half of the continuum eventually lead to a warping of the perceptual space in which within-category differences are reduced and between-category differences enlarged. Graphics inspired by Kuhl (2000).

4 | WHERE AND HOW ARE PERCEPTUAL CATEGORIES REPRESENTED?

There is no universal answer to the question *where* and *how* perceptual categories are represented in the brain. This is mainly due to the fact that the representation of categories is linked to both, sensory levels, as well as higher-order processes involved in the categorization task. Thus, forming new categories or assigning a new stimulus to an existing category requires the integration of bottom-up stimulus driven information with knowledge from prior experience and memory as well as linking this information to the appropriate response in case of an active categorization task. This complex and dynamic process is mediated by various neural structures and their relative contribution and interaction is currently still under investigation.

Categorical representations of sensory information must exhibit a high degree of flexibility to tolerate the variability across members of one category and provide at the same time perceptual stability. It is still unresolved at which processing level this perceptual invariance is implemented. In the auditory domain, three non-exclusive scenarios should be considered: (1) There are category-selective modules in the temporal

lobe, specialized for the abstraction level required for appropriate categorization, (2) low-level abstraction of acoustic information is achieved in the auditory cortex while a higher-level abstraction for action is accomplished only through dynamic interaction with categorical decision processes in prefrontal cortex (PFC), (3) the categorical percept arises from the invariant motor code used for the different sounds, in correspondence with the motor theory of speech perception (Liberman & Mattingly, 1985). While the latter scenario could only account for speech categories or action-related sounds, there is evidence for both low- as well as high-level abstraction processes in temporal and frontal cortex.

Categorization bridges the gap between lower-level sensory processing and higher-order cognition

Traditionally, higher-order sensory processing is divided into a ventral object-processing stream or 'what'-pathway and a dorsal spatial-processing stream also called 'where'-pathway, originally based on visual data (Ungerleider & Mishkin, 1982) and later confirmed also in the auditory domain (Romanski et al., 1999; Rauschecker & Tian, 2000). In accordance with this anatomical-functional differentiation, visual object processing was shown to rely on a progressively increasing complexity of the neural representations along the ventral stream from simple objects (Tanaka, 1993) to category-selective and highly invariant representations of faces in the fusiform face area (Kanwisher et al., 1997). Analogously, in the auditory domain, category-selective modules for voices, speech, and music have been found in the temporal lobe (Belin et al., 2000; Leaver & Rauschecker, 2010). Native phoneme categories were shown to specifically recruit the left superior temporal sulcus (STS) (Liebenthal et al., 2005) and the activation level of this region seems to correlate with the degree of categorical processing (Desai et al., 2008).

However, applications of modern fMRI analysis techniques revealed categorical representations in locally distributed and even overlapping activation patterns in the visual (Haxby et al., 2001) and auditory domain

(Staeren et al., 2009) challenging the strictly modular view. Staeren et al. (see also Leaver and Rauschecker, 2010) specifically addressed the natural correspondence of acoustic features and categorical groupings, which challenges previous reports of categorical representations, questioning whether these reflect true categories or 'merely' common physical attributes. Furthermore, pioneering visual category-learning studies from David Freedman and colleagues (Freedman et al., 2001, 2003) in monkeys revealed a discrepancy between the representation of low-level diagnostic stimulus features in the inferotemporal lobe and true categorical representations in PFC effectively used for the categorization task. This two-stage hierarchical model of categorization was supported by evidence from visual category learning studies with human subjects (Jiang et al., 2007; Li et al., 2009), and suggests that extrastriate[4] visual areas in the occipito-temporal lobe only build a preparatory stage for categorization in providing a task-independent representation of the category-relevant stimulus features that become available for read-out to higher-order processes in the frontal lobe for active categorization.

Recently, a similar dichotomy manifests in the auditory domain. On the one hand, imaging studies provided evidence for a generic role of bilateral STS in categorical processing beyond speech-specific categories (Leech et al., 2009; Liebenthal et al., 2010; Klein & Zatorre, 2011), on the other hand, carefully defined contrasts between sound discrimination and sound categorization tasks as well as specific fMRI adaptation designs have revealed the discrepancy between temporal and frontal categorical sound representations. Similarly to the results from visual experiments, highly abstract representations (i.e. invariant to changes within a phonetic category) seemed to rely on decision-related processes in the frontal lobe (Myers et al., 2009), which are also involved in active categorization of newly learned non-speech categories (Husain et al., 2006).

[4] Extrastriate cortex: visual areas in the occipito-temporal cortex, excluding primary visual cortex (V1 or striate cortex)

Taken as a whole, the body of literature suggests that categorical sound processing is accomplished in least two different levels of the auditory processing hierarchy; the specific role of category-selective modules and the level of abstraction from the acoustic features in the temporal lobe remain however elusive. This remaining ambiguity mainly stems from the prevalent use of speech sounds, for which humans are processing experts (similar to faces) even prior to linguistic experience (Eimas et al., 1987). Thus, results from categorical speech processing are not generalizable to other natural categories and especially lack insights into the processes underlying the formation of novel sound categories. A seminal study with gerbils demonstrated that learning to categorize artificial sounds in the form of frequency sweeps resulted in a transition from a physical to a categorical sound representation already in the primary auditory cortex (A1) (Ohl et al., 2001). In contrast to the traditional understanding of A1 as a feature detector, this finding implicates that sound representations at the first cortical analysis stage are malleable to environmental demands and prone to plastic reorganization. In fact, sound stimuli have passed through several levels of basic feature analyses before they ascend to the superior temporal cortex (Nelken, 2004). Thus, sound representations in A1 are unlikely to be faithful copies of the physical characteristics. It remains to be investigated whether the experience-driven abstract coding in early auditory cortex, observed in animals, is transferrable to humans. While the involvement of the auditory cortex (AC) in categorization of artificial sounds has also been demonstrated in humans (Guenther et al., 2002), these studies relied on conventional subtraction paradigms, lacking sufficient sensitivity to demarcate distinct categorical representations. Due to the large physical variability within categories and the similarity of sounds straddling the category boundary, between-category contrasts often do not reveal significant results (Klein & Zatorre, 2011). Modern analyses techniques with increased sensitivity to locally distributed activation changes in absence of changes in overall signal level (see Imaging and Analysis Methods) provide a promising tool to reveal perceptually invariant sound representations (Formisano et al., 2008; Kilian-Hütten et al., 2011).

5 | LEARNING AND PLASTICITY

The literature review given in the previous section revealed the methodological difficulties to separate categorical from acoustical sound representations as well as the challenge to investigate abstract sound representations without interfering with existing category modules for familiar sounds. The former has been tackled in the past with highly controlled stimulus designs, in which the acoustic differences between sound categories were minimized (Staeren et al., 2009). The latter can only be addressed with learning paradigms using novel sound stimuli. The complex and ever-changing natural acoustic environment requires the continuous adaptation of sound representations according to new experiences and requirements. While a detailed description of the learning process itself and the underlying mechanisms is beyond the scope of this thesis, this section represents an attempt to provide an overview of the major principles governing learning-induced neuroplasticity[5].

Learning can refer to early onset long-term exposure without explicit feedback or category labels as in the case of speech, but also to short-term adaptations to the demands of the current situation and task. While it has been demonstrated that mere exposure to stimuli can already facilitate their later discrimination without feedback-guided supervision (Goudbeek et al., 2009), the physical stimulus information alone is often not sufficiently informative about which features are relevant for categorization and which are not. This requires the integration of context and feedback in the formation of new categories. Research on the neural correlates of active category learning has revealed a multitude of brain regions, ranging from low-level sensory cortex related to stimulus processing over regions involved in attention, working-memory, decision-making as well as rule-learning in higher-order prefrontal regions to long-term memory, reward, strategy and feedback-related processes in medial temporal lobe structures

[5] Neuroplasticity: changes in neuronal organization (at the level of individual cells and their connections) as a result of maturation, learning, and pathology.

as well as stimulus-response associations and response preparation in premotor and motor areas. This (incomplete) list reveals that category learning is a very complex cognitive task that engages almost the whole brain and requires dynamic interactions between several different neural systems (for an elaborate description, please refer to Seger & Miller, 2010). Furthermore, there is considerable evidence that the contribution of the neural systems differs depending on the category-learning task (Ashby & Spiering, 2004). Typically, mainly three types are distinguished, namely rule-based tasks, information integration tasks and prototype distortion tasks, characterized by increasing levels of complexity. The experiments described in the course of this thesis all employ the simplest form of rule-based tasks where the categorization rule could in principle be described verbally as only one stimulus dimension is relevant for categorization.

Plasticity is an obligatory consequence of all neural activity

~ Pascual-Leone et al. (2005)

The processes and sites of neuroplasticity underlying category learning are still quite poorly understood. Hypotheses are primarily derived from perceptual learning studies in animals. These studies revealed that extensive discrimination training elicits reorganization of the auditory cortical maps, selectively increasing the representation of the behaviorally relevant sound features (Recanzone et al., 1993; Polley et al., 2006). This suggests that environmental and behavioral demands lead to changes of the auditory tuning properties of neurons such that more neurons are tuned to the relevant features to achieve higher sensitivity in the relevant dimension. This reorganization is mediated by synaptic plasticity, i.e. the strengthening of neuronal connections following rules of Hebbian learning[6].

[6] Hebbian learning: repeated co-activation of connected neurons induces synaptic strengthening. '*Cells that fire together, wire together.*' (Hebb, 1949; for recent review, see Caporale & Dan, 2008)

An important aspect of rule-based category learning is increased attention to the category-relevant features and the suppression of irrelevant features. The task therefore evokes an imbalance of attention that could induce selective reorganization of the cortical sound representations. As a consequence, more neural resources would be allocated to the behaviorally relevant information at the expense of information that is irrelevant for the decision. The adaptive allocation of neural resources to diagnostic information after category learning is supported by evidence from monkey electrophysiology (Sigala & Logothetis, 2002; De Baene et al., 2008) and human imaging, showing decreased activation for prototypical exemplars of a category relative to exemplars near the category boundary (Guenther et al., 2002). Reducing the representation of a large number of sounds to few relevant features presents an enormous processing advantage. It facilitates the read-out of the categorical pattern due to the reduced data structure and limits the neural resources by avoiding redundancies in the representation according to the concept of sparse coding (Olshausen & Field, 2004).

While passive learning studies suggest that attention is not necessary for sensory plasticity to occur (Watanabe et al., 2001; Seitz and Watanabe, 2003), the selective enhancement of features requires some top-down gating mechanism that specifies which features are behaviorally relevant and which are not. Attention can act as such a filter, selectively increasing feature saliency (Lakatos et al., 2013) by selectively modulating the tuning properties of neurons in the auditory cortex (Ahveninen et al., 2011) eventually leading to a competitive advantage of relevant information.

To date, the neural circuitry between sensory and higher-order attentional processes mediating learning-induced plasticity remain largely elusive. Predictive coding models propose that the dynamic interaction between bottom-up sensory information and top-down modulation by prior experience shapes the perceptual sound representation (Friston, 2005). This implies that categorical perception would arise from the continuous updating of the internal representation during learning to incorporate all variability present within a category, with the objective of reducing the prediction error (i.e. the difference between sensory input and internal

representation). Consequently, lasting interaction between forward driven processing and backward modulation could induce synaptic plasticity and result in an internal representation that correctly matches the categorical structure and therefore optimally guides correct behavior also beyond the scope of the training period. The implementation of these Bayesian processing models rests on fairly hierarchical structures consisting of forward, backward and lateral connections entering different cortical layers. According to the Reverse Hierarchy Theory (Ahissar & Hochstein, 2004), category learning would be initiated by high-level processes involved in rule-learning, controlling via top-down modulation selective plasticity at lower-level sensory areas sharpening the responses according to the learning rule. In accordance with this view, initial categorization performance was shown to correlate with prefrontal cortex activity, progressively decreasing with expertise (Little & Thulborn, 2005).

6 | MULTISENSORY INTERACTION AND CROSSMODAL PLASTICITY

Although category learning plays a fundamental role in natural adaptive behavior and the perception of natural environments is essentially multisensory (Stein & Meredith, 1993), little attention has been paid to the crossmodal influence on rapid learning-induced plasticity.

In the last years, the view that perception is largely driven by the interaction between the senses has gained acceptance. Integrating sensory signals from two modalities can lead to marked improvements in the detection and identification of stimuli by increasing the information content and optimization of the processing strategy, especially when the information from one modality is ambiguous or suboptimal (Ernst & Bülthoff, 2004). Under coherent circumstances (i.e. temporally and/or spatially corresponding), visual and auditory information complement each other, providing a more robust and reliable percept. Multisensory integration has mostly facilitative effects (Sumby & Pollack, 1954; Schürmann et al., 2004) however, the sensory system can be 'fooled' by mismatching signals, leading to illusory percepts. The McGurk effect (McGurk & MacDonald,

1976) is one of the classic examples of this phenomenon: Listening to a person uttering the phoneme /ba/ while watching a silent video of a person pronouncing the phoneme /ga/ results in the percept of a third phoneme /da/. Likewise, Jousmäki and Hari (1998) demonstrated a strong interaction between auditory and tactile stimuli in the 'parchment-skin illusion'. Their experiment revealed that a manipulation of the frequency content of the audio feedback induced by rubbing the palms together consistently modified the perceived roughness of the skin.

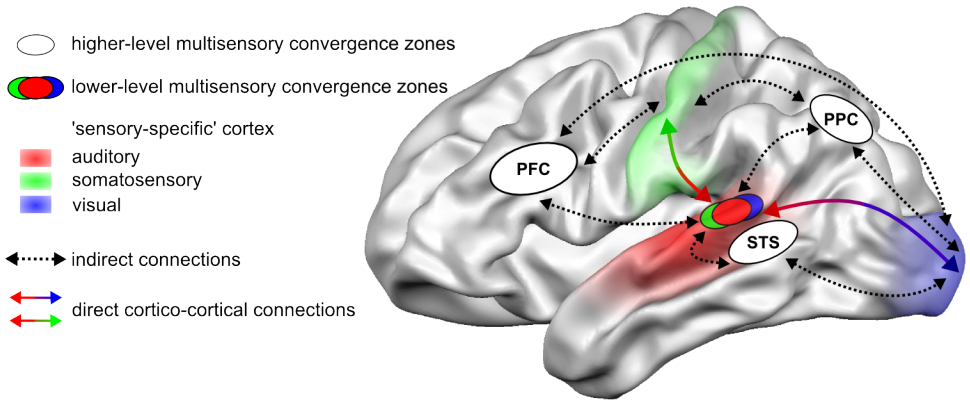


Figure 1.4 Simplified schematic representation of possible neural pathways mediating crossmodal interactions. Multisensory interaction is assumed to rely on higher-order regions in the posterior parietal cortex (PPC), inferior prefrontal cortex (PFC) as well as the superior temporal sulcus (STS). More recently, crossmodal activations have also been observed in classical 'sensory-specific' processing sites (here, in line with the focus of the thesis, only depicted for the auditory cortex). Despite evidence for crossmodal interactions in subcortical clusters (Stein & Meredith, 1993), thalamic-cortical projections are omitted in this overview. Adapted from Driver & Noesselt (2008).

Traditionally, merging of the sensory input from different modalities was assumed to take place in higher-order putative multisensory convergence zones such as the superior temporal sulcus (STS), the posterior parietal cortex (PPC) as well as the prefrontal cortex (PFC) after unisensory processing in 'sensory-specific' regions was complete (**Figure 1.4**, for

review, see Kaas & Collins, 2004). In the past decade, however, a growing number of studies have revealed anatomical connections between visual, auditory and somatosensory cortex (Falchier et al., 2002; Ro et al., 2009) building the direct route for crossmodal interactions at early processing stages (for a recent review, see Driver & Noesselt, 2008) challenging this strictly hierarchical model. Electrode recordings in the auditory cortex of monkeys further revealed the crucial dissociation between feedforward and feedback input patterns for visual and somatosensory stimuli. The laminar activation profile suggests that somatosensory input reaches the auditory cortex via both, forward and backward projections. In contrast, visual input in the auditory cortex seems to be confined to feedback from higher-order multisensory regions (Schroeder & Foxe, 2002). The functional dissociation between higher-level and lower-level multisensory integration remains an important issue for future research.

Crucially, animal research indicates that the mechanisms of multisensory integration are not hard-wired at birth but rather formed through experience. Cats reared in altered sensory environments established principles of multisensory integration that are absent during normal development (Wallace & Stein, 2007). These findings are complemented by reports of crossmodal plasticity in humans suffering from sensory deprivation. The loss of a sensory modality is often compensated by heightened sensitivity in the remaining modalities (Lessard et al., 1998; Gougoux et al., 2004, 2005) and recent experimental and clinical findings have revealed that the source of these superior abilities is reorganization of the sensory processing mechanisms (Hamilton et al., 2000; Lomber et al., 2010). Especially in the early blind or early deaf, increased recruitment of the deprived sensory system by the remaining senses has been observed. A common example is the primary visual cortex activation in Braille reading blind subjects (Sadato et al., 1996). The similarity of auditory and vibrotactile stimuli probably underlies the frequency-specific responses in the auditory cortex induced by tactile stimulation in deaf subjects (Levänen et al., 1998). While this corroborates the functional relevance of crossmodal plasticity after sensory deprivation, processing of ecologically meaningful associations such as

voices and faces seems to benefit from the same crossmodal mechanism in normally functioning subjects. Watching a silent video of a speaking person was shown to recruit brain regions central to auditory processing including Heschl's gyrus and its adjacency (Calvert et al., 1997; Pekkola et al., 2005). Likewise, processing familiar voices engaged face-specific areas along the higher-order visual processing stream in absence of visual input (Von Kriegstein et al., 2005).

Remarkably, already short-term manipulations of visual input through blindfolding can induce crossmodal activity in the primary visual cortex by tactile stimuli, accompanied by superior Braille reading performance (Pascual-Leone & Hamilton, 2001; Kauffman et al., 2002). In comparison to the extensive crossmodal plasticity in congenitally blind subjects that is manifested in new and strengthened connections between sensory cortices developing in the course of long-term deprivation, rapidly induced crossmodal effects are reversible and most likely promoted by unmasking existing connections (Pascual-Leone et al., 2005). Although the degree of neural plasticity induced by learning and experience in the adult brain is minor compared to development, it has been shown that multimodal training designs are generally superior to unimodal training designs (Shams & Seitz, 2008). The beneficial effect of multisensory exposure during training even lasts beyond the training period itself reflected in increased performance after removal of the stimulus from one modality (for review, see Shams, Wozny, Kim, & Seitz, 2011). This effect has been demonstrated even for brief training periods and arbitrary stimulus pairs (Ernst, 2007), promoting the view that short-term multisensory learning can lead to lasting reorganization of the processing networks.

Compared to the large body of literature on crossmodal plasticity during development or in case of sensory loss, research on rapid, learning-induced crossmodal plasticity in the mature brain is scarce. Given that category learning relies to a large degree on the integration of contextual information it can be assumed that there is enormous crossmodal influence during the abstraction process. Therefore, it is of major interest to explore how

qualitatively different but complementary information from two modalities shape the representations of sounds. Particularly, it remains to be investigated whether the similarities of auditory and tactile stimuli in terms of vibratory information content provide a processing advantage in early sensory cortex mediating plasticity at different (potentially earlier) levels than stimuli from the auditory and visual modalities. Hitherto, insights from comparative learning studies in this respect are lacking.

7 | IMAGING AND ANALYSIS METHODS

Most brain imaging studies presented in this thesis employed fMRI, a non-invasive measure indirectly sensitive to local changes in blood oxygenation, which are coupled to the metabolic rate of neurons. Active neurons consume energy and they extract oxygen supplied in the blood to compensate for this disparity. To meet the increased demand, blood flow to the respective active brain regions rapidly increases. Due to the different magnetic properties of oxygenated and deoxygenated hemoglobin[7], the dynamics of the metabolism and blood flow can be visualized as they differentially affect the magnetic signal emitted from hydrogen nuclei in the body. The response measured with fMRI is henceforth a blood-oxygen-level-dependent (BOLD) contrast, whose timing depends on the velocity of blood flow, which peaks after approximately six seconds following the event that caused the neurons to respond before if relapses back to baseline level.

Conventional (i.e. univariate) fMRI data analyses rely on voxel-wise[8] subtraction of mean activation levels for different experimental conditions (or one experimental condition and a rest condition) to reveal those brain areas specifically engaged in certain sensory or cognitive processes. To this end, the variability of a voxel's activation level over the time course of the experiment is modeled by the linear combination of its predicted responses

[7] Hemoglobin is the iron-containing protein in the red blood cells that carries oxygen
[8] A voxel is the three-dimensional equivalent of a pixel

to the different experimental conditions. How well a certain condition explains the voxel's activation change is represented in the model coefficients or beta-weights. Finally, to estimate the relative preference of a voxel for one condition versus another, contrasts between beta-weights can be formulated and the difference can be tested statistically.

The studies presented in this thesis particularly aimed at revealing the changes in sound representations underlying the discrepancy between physical and perceptual similarity induced by category learning. In other words, the experiments were designed to identify those brain regions that specifically code the arbitrary perceptual separation of sound categories despite uniform distribution of the acoustic information. Different perceptual states may however not necessarily be reflected in differences of a voxel's activation level (Ohl et al., 2001; Schnupp et al., 2006). In that case the univariate analysis of the between-category contrast would not reveal any significant response differences. In contrast, evidence suggests that categories are coded in the combined information of multiple locally distributed voxels (Haxby et al., 2001; Staeren et al., 2009). In the past decade, machine-learning algorithms have advanced into the field of neuroscientific research, expanding the possibilities of multivariate analysis techniques to capture these spatially distributed effects otherwise inscrutable to traditional univariate analyses. Therefore, in addition to conventional univariate analyses, the studies presented in the following chapters employed an approach called multi-voxel pattern analysis (MVPA, see Haxby, 2012; Haynes & Rees, 2006; Norman, Polyn, Detre, & Haxby, 2006 for recent tutorial reviews). In the analyses presented in this thesis, linear Support Vector Machines (SVM, Cox & Savoy, 2003) were used as classification algorithms that learn by means of a number of so-called 'training trials' to distinguish spatial activation patterns corresponding to the joint activity of a number of voxels based on the different experimental conditions or perceptual states. After successful training, the established model can be used to predict for every new trial the corresponding

perceptual state or condition. The training and testing procedure is explained in more detail in the respective chapters (chapter 2 and 4).

8 | THESIS OUTLINE

The background information provided in this chapter has revealed the vital role of experience and learning in shaping sound representations to comply with the environmental demands. This also includes the integration of cues from different modalities to achieve a robust and behaviorally relevant representation. Reviewing the literature on categorical sound representations also illustrated the challenges to separate physical from perceptual similarity imposed by the large variability within abstract sound categories. The experimental studies presented in the following chapters combat these challenges with the combination of a carefully designed psychophysical approach and advanced fMRI analyses techniques to reveal learning-induced changes in sound representations under constant acoustic stimulation conditions. Using novel artificial sound categories, the study presented in **chapter 2**, directly compares pre- and post-training sensitivity of neural populations in the auditory cortex to between-category contrasts. Crucially, categorical representations are assessed with a passive listening design to avoid interference with memory and decision-related processes. This study aims at elucidating the plastic nature of early auditory cortex in the context of abstraction learning. **Chapter 3** expands the category learning design to incorporate multisensory cues rather than explicit category labels in accordance with natural learning conditions. In a unique comparison of matched audio-tactile and audio-visual association training conditions, this study investigates learning-induced crossmodal reorganization of sound responses. In **chapter 4** both approaches are combined to investigate the functional role of sound representations formed through crossmodal association training. Specifically, this chapter aims to reveal potential differences in the categorical pitch representations after audio-visual and audio-tactile training. In **chapter 5**, the thesis concludes with a summary of

the main experimental findings and briefly addresses the relevance of this research for clinical applications.

REFERENCES

Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. Trends Cogn Sci 8: 457–464.

Ahveninen J, Hämäläinen M, Jääskeläinen IP, Ahlfors SP, Huang S, Lin F-H, Raij T, Sams M, Vasios CE, Belliveau JW (2011) Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. Proc Natl Acad Sci U S A 108: 4182–4187.

Ashby FG, Spiering BJ (2004) The neurobiology of category learning. Behav Cogn Neurosci Rev 3: 101–113.

De Baene W, Ons B, Wagemans J, Vogels R (2008) Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. Lern Mem 15: 717–727.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403: 309–312.

Benders T, Escudero P, Sjerps M (2010) The interrelaton between acoustic context effects and available response categories in speech sound categorization. J Acoust Soc Am 131: 3079–3087.

Benzeguiba M, Mori R De, Deroo O, Dupont S, Erbes T, Jouvet D, Fissore L, Laface P, Mertins A, Ris C, Rose R, Tyagi V, Wellekens C (2007) Automatic speech recognition and speech variability: a review. Speech Commun 49: 10–11.

Bornstein MH (1987) Perceptual Categories in Vision and Audition. In: Categorical Perception. The Groundwork of Cognition (Harnad S, ed), pp.287–300. Cambridge, UK: Cambridge UP.

Bornstein MH, Kessen W, Weiskopf S (1976) Color vision and hue categorization in young human infants. J Exp Psychol Hum Percept Perform 2: 115–129.

Brown AM, Lindsey DT, Guckes KM (2011) Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. J Vis 11: 1–21.

Burns EM, Ward WD (1978) Categorical perception-phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. J Acoust Soc Am 63: 456–468.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. Science 276: 593–596.

Caporale N, Dan Y (2008) Spike timing-dependent plasticity: a Hebbian learning rule. Annu Rev Neurosci 31: 25–46.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19: 261–270.

Desai R, Liebenthal E, Waldron E, Binder JR (2008) Left posterior temporal regions are sensitive to auditory categorization. J Cogn Neurosci 20: 1174–1188.

Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. Neuron 57: 11–23.

Eggermont JJ (2001) Between sound and perception: reviewing the search for a neural code. Hear Res 157: 1–42.

Eimas PD, Miller JL, Jusczyk PW (1987) On infant speech perception and the acquisition of language. In: Categorical Perception. The Groundwork of Cognition (Harnad S, ed), pp.161–195. Cambridge, Massachusetts: Cambridge UP.

Ernst MO (2007) Learning to integrate arbitrary signals from vision and touch. J Vis 7: 1–14.

Ernst MO, Bülthoff HH (2004) Merging the senses into a robust percept. Trends Cogn Sci 8: 162–169.

Etcoff NL, Magee JJ (1992) Categorical perception of facial expressions. Cognition 44: 227–240.

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22: 5749–5759.

Fodor JA (1975) The Language of Thought. New York: Crowell.

Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322: 970–973.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291: 312–316.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J Neurosci 23: 5235–5246.

Friston K (2005) A theory of cortical responses. Phil Trans R Soc B 360: 815–836.

Galantucci B, Fowler CA, Turvey MT (2006) The motor theory of speech perception reviewed. Psychon Bull Rev 13: 361–377.

Gilbert AL, Regier T, Kay P, Ivry RB (2006) Whorf hypothesis is supported in the right visual field but not the left. Proc Natl Acad Sci U S A 103: 489–494.

Goldstone R (1994) Influences of categorization on perceptual discrimination. J Exp Psychol Gen 123: 178–200.

Goudbeek M, Swingley D, Smits R (2009) Supervised and unsupervised learning of multidimensional acoustic categories. J Exp Psychol Hum Percept Perform 35: 1913–1933.

Gougoux F, Lepore F, Lassonde M, Voss P, Zatorre RJ, Belin P (2004) Pitch discrimination in the early blind. Nature 430.

Gougoux F, Zatorre RJ, Lassonde M, Voss P, Lepore F (2005) A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals. PLoS Biol 3: e27.

Guenther FH, Husain FT, Cohen MA, Shinn-Cunningham BG (1999) Effects of categorization and discrimination training on auditory perceptual space. J Acoust Soc Am 106: 2900–2912.

Guenther FH, Nieto-Castanon A, Ghosh SS, Tourville JA (2002) Representation of sound categories in auditory cortical maps. J Speech Lang Hear Res 47: 46–57.

Hamilton R, Keenan JP, Catala M, Pascual-Leone a (2000) Alexia for Braille following bilateral occipital stroke in an early blind woman. Neuroreport 11: 237–240.

Harnad S (1987) Categorical perception: The groundwork of cognition (Harnad S, ed). Cambridge, UK: Cambridge UP.

Haxby JV (2012) Multivariate pattern analysis of fMRI: the early beginnings. NeuroImage 62: 852–855.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293: 2425–2430.

Haynes J-D, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7: 523–534.

Hebb DO (1949) The First Stage of Perception: Growth of the Assembly. In: The Organization of Behavior: A Neuropsychological Theory, pp.49–56. New York: Wiley.

Husain FT, Fromm SJ, Pursley RH, Hosey L, Braun A, Horwitz B (2006) Neural bases of categorization of simple speech and nonspeech sounds. Hum Brain Mapp 27: 636–651.

Jiang X, Bradley E, Rini R a, Zeffiro T, Vanmeter J, Riesenhuber M (2007) Categorization training results in shape- and category-selective human neural plasticity. Neuron 53: 891–903.

Jousmäki V, Hari R (1998) Parchment-skin illusion: sound-biased touch. Curr Biol 8: 190–191.

Kaas JH, Collins CE (2004) The Resurrection of the Multisensory Cortex. In: The Handbook of Multisensory Processes (Calvert GA, Spence C, Stein BE, eds), pp.285–294. Cambridge, Massachusetts: Bradford.

Kandel ER, Schwartz JH, Jessel TM eds. (2000) Principles of Neural Science. In. New York: Mcgraw-Hill Professional.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17: 4302–4311.

Kauffman T, Théoret H, Pascual-Leone A (2002) Braille character discrimination in blindfolded human subjects. Neuroreport 13: 571–574.

Kilian-Hütten N, Valente G, Vroomen J, Formisano E (2011) Auditory cortex encodes the perceptual interpretation of ambiguous sound. J Neurosci 31: 1715–1720.

Klein ME, Zatorre RJ (2011) A role for the right superior temporal sulcus in categorical perception of musical chords. Neuropsychologia 49: 878–887.

Kohler E, Keysers C, Umiltà MA, Fogassi L, Gallese V, Rizzolatti G (2002) Hearing sounds, understanding actions: action representation in mirror neurons. Science 297: 846–848.

Kowalski N, Depireux DA, Shamma SA (1996a) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J Neurophysiol 76: 3503–3523.

Kowalski N, Depireux DA, Shamma SA (1996b) Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. J Neurophysiol 76: 3524–3534.

Von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud A-L (2005) Interaction of face and voice areas during speaker recognition. J Cogn Neurosci 17: 367–376.

Kuhl PK (1991) Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. Percept Psychophys 50: 93–107.

Kuhl PK (2000) A new view of language acquisition. Proc Natl Acad Sci U S A 97: 11850–11857.

Kuhl PK, Miller JD (1975) Speech perception by the chinchilla: voiced-voiceless distrinction in alveolar plosive consonants. Science 190: 69–72.

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77: 750–761.

Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci 30: 7604–7612.

Leech R, Holt LL, Devlin JT, Dick F (2009) Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. J Neurosci 29: 5234–5239.

Lessard N, Paré M, Lepore F, Lassonde M (1998) Early-blind human subjects localize sound sources better than sighted subjects. Nature 395: 278–280.

Levänen S, Jousmäki V, Hari R (1998) Vibration-induced auditory-cortex activation in a congenitally deaf adult. Curr Biol 8: 869–872.

Li S, Mayhew SD, Kourtzi Z (2009) Learning shapes the representation of behavioral choice in the human brain. Neuron 62: 441–452.

Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. Cognition 21: 1–36.

Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol 54: 358–368.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. Cereb Cortex 15: 1621–1631.

Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR (2010) Specialization along the left superior temporal sulcus for auditory categorization. Cereb Cortex 20: 2958–2970.

Little DM, Thulborn KR (2005) Correlations of cortical activation and behavior during the application of newly learned categories. Cogn Brain Res 25: 33–47.

Livingston KR, Andrews JK, Harnad S (1998) Categorical perception effects induced by category learning. J Exp Psychol Learn Mem Cog 24: 732–753.

Lomber SG, Meredith MA, Kral A (2010) Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. Nat Neurosci 13: 1421–1427.

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264: 746–748.

Minshew NJ, Meyer J, Goldstein G (2002) Abstract reasoning in autism: a disassociation between concept formation and concept identification. Neuropsychology 16: 327–334.

Myers EB, Blumstein SE, Walsh E, Eliassen J (2009) Inferior frontal regions underlie the perception of phonetic category invariance. Psychol Sci 20: 895–903.

Myers EB, Swan K (2012) Effects of category learning on neural sensitivity to non-native phonetic categories. J Cogn Neurosci 24: 1695–1708.

Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14: 474–480.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10: 424–430.

Ohl FW, Scheich H, Freeman WJ (2001) Change in pattern of ongoing cortical activity with auditory category learning. Nature 412: 733–736.

Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. Curr Opin Neurobiol 14: 481–487.

Pascual-Leone A, Amedi A, Fregni F, Merabet LB (2005) The plastic human brain cortex. Annu Rev Neurosci 28: 377–401.

Pascual-Leone A, Hamilton R (2001) The metamodal organization of the brain. Prog Brain Res 134: 427–445.

Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, Sams M (2005) Primary auditory cortex activation by visual speech: an fMRI study at 3T. Neuroreport 16: 125–128.

Polley DB, Steinberg EE, Merzenich MM (2006) Perceptual learning directs auditory cortical map reorganization through top-down influences. J Neurosci 26: 4970–4982.

Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R (2009) Neural correlates of categorical perception in learned vocal communication. Nat Neurosci 12: 221–228.

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A 97: 11800–11806.

Recanzone GH, Schreiner CE, Merzenich MM (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J Neurosci 13: 87–103.

Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annu Rev Neurosci 27: 169–192.

Ro T, Hsu J, Yasar NE, Elmore LC, Beauchamp MS (2009) Sound enhances touch perception. Exp Brain Res 195: 135–143.

Roberson D, Davies I, Davidoff J (2000) Color categories are not universal: replications and new evidence from a stone-age culture. J Exp Psychol Gen 129: 369–398.

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2: 1131–1136.

Rosch E (1998) Principles of Categorization. In: Concepts: Core Readings (Margolis E, Laurence S, eds). Cambridge, Massachusetts: MIT Press.

Sadato N, Pascual-Leone A, Grafman J, Ibanez V, Deiber M, Dold G, Hallett M (1996) Activation of the primary visual cortex by Braille reading in blind subjects. Nature 380: 526–528.

Schnupp JWH, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J Neurosci 26: 4785–4795.

Schönwiesner M, Zatorre RJ (2009) Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proc Natl Acad Sci U S A 106: 14611–14616.

Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cogn Brain Res 14: 187–198.

Schürmann M, Caetano G, Jousmäki V, Hari R (2004) Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. J Acoust Soc Am 115: 830.

Seger CA, Miller EK (2010) Category learning in the brain. Annu Rev Neurosci 33: 203–219.

Seitz AR, Watanabe T (2003) Is subliminal learning really passive? Nature 422: 2003.

Shams L, Seitz AR (2008) Benefits of multisensory learning. Trends Cogn Sci 12: 411–417.

Shams L, Wozny DR, Kim R, Seitz A (2011) Influences of multisensory experience on subsequent unisensory processing. Front Psychol 2: 264.

Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415: 318–320.

Sridharan M, Stone P (2009) Color learning and illumination invariance on mobile robots: A survey. Robot Auton Syst 57: 629–644.

Staeren N, Renvall H, De Martino F, Goebel R, Formisano E (2009) Sound categories are represented as distributed patterns in the human auditory cortex. Curr Biol 19: 498–502.

Stein BE, Meredith MA (1993) The Merging of the Senses. Cambridge, Massachusetts: MIT Press.

Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. J Acoust Soc Am 26: 212–215.

Tanaka K (1993) Neuronal Mechanisms of Object Recognition. Science 262: 685–688.

Ungerleider LG, Mishkin M (1982) Two Cortical Visual Systems. In: Analysis of Visual Behavior (Ingle DJ, Goodale MA, Mansfield, Richard JW, eds), pp.549–586. Cambridge, Massachusetts: The MIT Press.

Wallace MT, Stein BE (2007) Early experience determines how the senses will interact. J Neurophysiol 97: 921–926.

Watanabe T, Náñez JE, Sasaki Y (2001) Perceptual learning without perception. Nature 413: 844–848.

Wyttenbach RA, May ML, Hoy, Ronald R (1996) Categorical perception of sound frequency by crickets. Science 273: 1542–1544.

Chapter 2

LEARNING OF NEW SOUND CATEGORIES SHAPES NEURAL RESPONSE PATTERNS IN HUMAN AUDITORY CORTEX

Ley, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., & Formisano, E. (2012). Learning of new sound categories shapes neural response patterns in human auditory cortex. J Neurosci 32: 13273-13280.

Every man can, if he so desires,
become the sculptor of his own brain

~ Santiago Ramón y Cajal

ABSTRACT

The formation of new sound categories is fundamental to everyday goal-directed behavior. Categorization requires the abstraction of discrete classes from continuous physical features as required by context and task. Electrophysiology in animals has shown that learning to categorize novel sounds alters their spatio-temporal neural representation at the level of early auditory cortex. However, functional magnetic resonance imaging (fMRI) studies so far did not yield insight into the effects of category learning on sound representations in human auditory cortex. This may be due to the use of overlearned speech-like categories and fMRI subtraction paradigms, leading to insufficient sensitivity to distinguish the responses to learning-induced, novel sound categories. Here, we used fMRI pattern analysis to investigate changes in human auditory cortical response patterns induced by category learning. We created complex novel sound categories and analyzed distributed activation patterns during passive listening to a sound continuum before and after category learning. We show that only *after* training, sound categories could be successfully decoded from early auditory areas and that learning-induced pattern changes were specific to the category-distinctive sound feature (i.e., pitch). Notably, the similarity between fMRI response patterns for the sound continuum mirrored the sigmoid shape of the behavioral category identification function. Our results indicate that perceptual representations of novel sound categories emerge from neural changes at early levels of the human auditory processing hierarchy.

1 | INTRODUCTION

Categorical Perception (CP) refers to the discrepancy between perceptual similarity and physical similarity of stimuli when they are grouped into distinct but meaningful classes (Harnad, 1987). Depending on situation and task, the relevant feature(s) defining the classes might differ. In the course of minimizing within-category and maximizing between-category differences, continuous physical variations between stimuli are overruled such that seemingly dissimilar stimuli may be considered 'same'. In audition, these perceptual transformations likely result in more abstract representations of sound similarity. Several attempts have been made to identify the neural source of these perceptual changes, however, to date the effects of category learning on sound representations could not be resolved in humans. Previous fMRI studies have relied on subtraction paradigms lacking sufficient sensitivity to distinguish the responses to novel sound categories and allowing only indirect inferences about the underlying changes in representation (Desai et al., 2008; Leech et al., 2009; Liebenthal et al., 2010). Furthermore, the use of speech-like sounds might obstruct the emergence of novel learning-induced category representations due to interference with existing phoneme representations.

In the visual domain, category learning is traditionally assumed to involve at least two different encoding stages: Whereas areas in the inferior temporal cortex are engaged in stimulus specific processes such as feature extraction, activation in the prefrontal cortex (PFC) codes more abstract, categorical information (Freedman et al., 2001, 2003; Seger & Miller, 2010). In contrast, animal electrophysiology in the auditory domain suggests that categorical sound information is encoded in spatio-temporal variations of neural firing already in early auditory cortex (Ohl et al., 2001; Selezneva et al., 2006). These changes in firing patterns might not necessarily lead to increases in overall activation level (Ohl et al., 2001; Schnupp et al., 2006). It has been proposed that multi-voxel pattern analysis (MVPA) is sensitive to changes in distributed activation patterns in absence of changes in overall activation level (Haxby et al., 2001). This method has been successfully

employed to reveal subtle differences in overlapping sound representations (Formisano et al., 2008; Staeren et al., 2009) and purely perceptual processes in the visual (Li et al., 2007, 2009) and auditory (Kilian-Hütten et al., 2011) domain.

Here, we used fMRI and MVPA techniques in combination with a recursive feature elimination (RFE) procedure (De Martino et al., 2008) to reveal changes in sound representations in human auditory cortex induced by the formation of new sound categories. Our sound categories comprised complex moving ripples (Kowalski et al., 1996a, 1996b) which share important spectro-temporal properties with natural sounds but cannot be associated with any pre-existing category. Using novel auditory categories, we avoided any confounding effects related to prior experience. We trained subjects to categorize the sounds into two distinct classes and measured fMRI responses to new sounds from the same categories before and after successful category learning during passive listening. We then aimed to decode the novel perceptual sound categories from the auditory response patterns in absence of an active categorization task.

2 | MATERIAL AND METHODS

Participants. Eight volunteers (3 males, mean age: 23.38 years) participated in the study after providing informed consent. Participants reported normal hearing abilities and were naïve to sounds and research question. The study procedures were approved by the local ethics committee (Ethische Commissie Psychologie at Maastricht University).

Stimuli. Ripple stimuli (see **Figure 2.1A**) have successfully been employed in the past for characterizing spectro-temporal response fields in animals and humans (Kowalski et al., 1996a, 1996b; Shamma, 1996; Depireux et al., 2001; Langers et al., 2003). Here, ripples were composed of 50 sinusoids with logarithmically spaced frequencies spanning four octaves. The lowest frequency component (f_0) of the complex was shifted between on average 168 - 236 Hz to modulate ripple pitch. To create different ripple

densities, their spectral envelope was modulated sinusoidally along the frequency axis on a linear amplitude scale by 0.25 and 0.125 cycles/octave. Additionally, a constant envelope drift along the frequency axis was introduced by shifting the phase of the sinusoid over time. The angular velocity of this drift was varied in equal steps between 1 and 6 cycles/second. Drift direction was downward with an initial phase of 0. The stimuli were of 1-second duration and their energy was matched by adjusting their root mean square (rms) values. Linear amplitude ramps of 5 ms duration were added at ripple on- and offsets. All stimuli were sampled at 44.1 kHz using 16-bit resolution and processed in Matlab (MathWorks).

Stimulus calibration. As previous experiments have shown that inter-subject differences in stimulus discrimination ability can be rather large (Guenther et al., 1999), participants underwent a short calibration procedure where pitch and velocity discrimination sensitivity of the ripple sounds used for category learning were measured to match task difficulty. For this purpose, an adaptive up-down staircase procedure (AX same-different paradigm) was employed. Following the procedure devised by Levitt (Wetherill and Levitt, 1965; Levitt, 1971), we estimated a Just Noticeable Difference (JND) at a probability of 71% 'different' responses at convergence based on 15 response reversals. Participants were exposed to a sequence of three sounds, which consisted of two ripple sounds (A and X) separated by a noise burst. The participants were instructed to compare the two ripple sounds and ignore the noise burst, which could be considered a 'masker' as it was introduced to interfere with the sensory trace of A and to promote the transformation of the feature-based representations into a categorical percept (Guenther et al., 1999). Importantly, the noise burst did not disrupt the perception of the preceding and following ripple sound. All sound features except the relevant one were kept constant during the calibration procedure. The pitch discrimination threshold measured around the category boundary served as a global estimate for the small range of frequencies used in the experiment. The average JND of ripple pitch [baseline value (f_0): 200 Hz] was 21.76 Hz (*SEM* = 3.51). The average JND

for velocity [baseline value: 1 cycle/second] was 0.21 cycles/second (*SEM* = 0.04), which was well below the step size of 1 cycle/second utilized in the construction of the sound categories. We therefore assume that the velocity differences in the sounds are sufficiently salient.

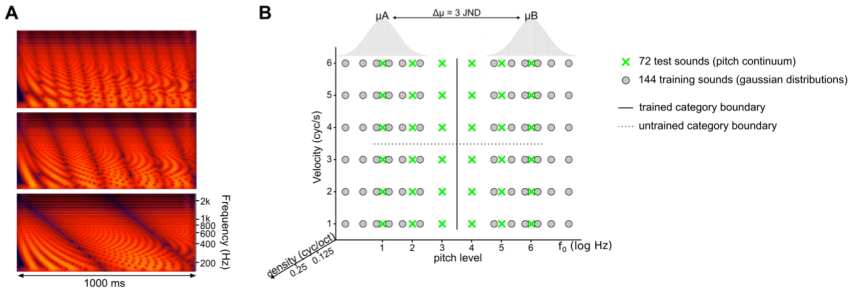


Figure 2.1 Sound spectrograms and stimulus space. A, Three example spectrograms of moving ripples with low (bottom), medium (middle) and high (upper) velocities at constant pitch and density values. **B**, Multidimensional stimulus space spanning the two categories A and B. The third dimension (density) is only partially indicated for clarity reasons. Similar to previous studies (Smits et al., 2006), pitch categories were defined by two non-overlapping one-dimensional Gaussian probability density functions (pdfs) on a logarithmic frequency scale. The distance between category means (μ_A and μ_B) was determined by individual psychometric measures (see Stimulus calibration) to match task difficulty. The category boundary was fixed at 200 Hz (f); standard deviations (σ) were set to one JND. During training, pdfs were linearly sampled resulting in two distinct pitch clusters containing six different values each (grey circles). In line with former behavioral studies on category learning (Smits et al., 2006; Goudbeek et al., 2009) six novel equidistant pitch values lying on a psychophysical continuum between category means were employed for scanning and to assess categorization performance outside the scanner (green crosses). Each pitch exemplar was presented with six different velocity and two different density values.

Category distributions. To partition a continuous stimulus space **(Figure 2.1B**) into different categories we used a combination of several spectral and temporal features (pitch, velocity, and density). We employed two distinct sets of sounds for category training and testing. Category training was restricted to one dimension (i.e. 'low pitch' vs. 'high pitch'). The additional spectral and temporal variations were introduced to encourage the

extraction of the category-distinctive sound feature under variable stimulus conditions and to promote the abstraction across task-irrelevant features. Categories were named A and B to avoid any explicit cues about the relevant sound feature. Instead, learning of the two pitch categories was encouraged by means of distributional information: For training, pitch values were sampled from two non-overlapping normal distributions with equal variance but different means defined on a logarithmic frequency scale (Equivalent Rectangular Bandwidth (ERB), Glasberg & Moore, 1990). Sampling was denser within categories than at the category border (see grey circles in **Figure 2.1B**). In contrast to pitch, the irrelevant dimensions (velocity and density) were linearly sampled. For fMRI sessions and to assess categorization performance in behavioral sessions outside the scanner, we used new test sounds (green crosses in **Figure 2.1B**). Crucially, these test sounds were evenly sampled from a psychophysical continuum between category means and therefore conveyed no information about the category boundary in terms of acoustic similarity. Due to the lack of distributional information, the test stimulus space was defined by equal variance in the relevant (i.e. pitch) as well as one of the irrelevant (i.e. velocity) dimensions and therefore allowed two equally feasible category partitions (see trained and untrained category boundary in **Figure 2.1B**). The division into two 'untrained' velocity classes ('slow' vs. 'fast') served as a control for the behavioral relevance of our results during the fMRI analysis (see below).

Experimental procedure. To ensure compatibility of sound quality during behavioral training and scanning, stimulus calibration and category training were performed inside the scanner room with the same hardware and audio settings as used during fMR imaging. Participants were seated on the scanner bed in comfortable viewing distance from the screen.

During behavioral sessions, training and test blocks were interleaved. The latter served to obtain consecutive measures of categorization performance and monitor the level of CP. For this purpose, we adapted a standard procedure from speech research (Liberman et al., 1957) in which subjects labeled the test sounds from the continuum without corrective

feedback. Participants always started with a test block, in which they were instructed to group the 72 sounds into two discrete classes (A vs. B) in a two-alternative forced choice procedure without instructions about the relevant stimulus dimension. The test block was followed by a training block comprising 144 sounds from the normal distributions. During training, visual feedback was provided after each response by means of a small red (incorrect) or green (correct) square appearing for 700 ms in the screen center. One training block lasted 12 minutes and allowed a short break after half of the trails. A test block lasted 6 minutes and was completed in one run. The number of repetitions and thereby the length of a behavioral training session was determined by the performance level (successful learning was determined by at least 85 % correct in one of the test blocks) as well as the motivation and condition of the participant, however, never exceeded one hour.

We measured fMRI responses to the 72 test sounds before and after successful category learning during passive listening (see *Imaging*). The first scan session was followed by a variable number (3 - 7) of behavioral training blocks, spread over 2 - 4 days so as to match subjects' performance before the second scanning session.

Curve fitting. We employed a curve fitting procedure (using Matlab's 'fit' function) to describe the learning-induced changes in sound labeling. Previous research (McMurray and Spivey, 2000) has shown that the s-shaped identification function in CP experiments resembles the logistic function, given by **(1)**.

$$y = \frac{a}{1 + e^{\frac{-(x-d)}{c}}} + b \qquad \textbf{(1)}$$

Here, *a* provides a measure of the amplitude of the function, *b* corresponds to the y-axis location of the lower asymptote, *c* reflects the slope of the function, and *d* indicates the location of the category boundary on the x-axis. We fitted the logistic function to the individual category identification

functions. The nonlinear least squares parameter estimation was subject to the following constraints: $0 \leq a \leq 100$; $0 \leq b \leq 100$; $0.1 \leq c \leq 10$; $1 \leq d \leq 6$. The liberal parameter settings were chosen to achieve a good fit and thereby provide an accurate description of the curve's shape and the underlying trend in the response data.

Imaging. Brain imaging was performed with a 3-Tesla Siemens Allegra MR headscanner at the Maastricht Brain Imaging Center (MBIC). For each subject, there were two scanning sessions, one pre- and the other post-category learning. In both these sessions, three runs (each consisting of 364 volumes and including the 72 test sounds; total number of sounds: 72 x 3 = 216) of functional MRI data were acquired in 30 slices, covering the temporal and parts of the frontal lobe with an eight channel head coil using a standard echo-planar imaging (EPI) sequence in a slow event-related design with the following parameters: Repetition time (TR) = 3500 ms; acquisition time (TA) = 2100 ms; field of view (FoV) = 224 x 224 mm; matrix size = 112 x 112; echo time (TE) = 30 ms; voxel dimensions = 2 mm isotropic. Additionally, anatomical T1-weighted images (voxel dimensions = 1 mm isotropic) were acquired with optimal grey-white matter contrast for cortex reconstruction purposes. The average inter-trial-interval between two stimuli was 17.5 s (jittered between 4, 5, and 6 TR). Sounds were delivered binaurally via MRI-compatible headphones (Visual Stim Digital, Resonance Technology Inc. or Sensimetrics S14, Sensimetrics Corporation) in the 1400 ms silent gaps between volume acquisitions. Stimulus order was randomized using the randperm function implemented in Matlab; stimulus delivery was synchronized with MR pulses using Presentation software (Neurobehavioralsystems).

FMRI preprocessing and univariate analysis. MRI data were first analyzed with BrainVoyager QX (Brain Innovations, Maastricht, The Netherlands). The first four volumes per run were discarded from the analysis to allow for T1 equilibrium. Functional data preprocessing included 3-dimensional head motion correction, slice scan-time correction (using sinc interpolation),

temporal high-pass filtering (3 cycles), linear trend removal, co-registration to individual structural images, and normalization of anatomical and functional data to Talairach space. Individual cortical surfaces were reconstructed from grey-white matter segmentations and aligned using a moving target-group average approach based on curvature information (cortex-based alignment, Goebel et al., 2006) to obtain an average 3D surface representation. For univariate statistical analysis of the functional data, a general linear model (GLM) was computed by fitting the blood-oxygen-level-dependent (BOLD) response time course with the predicted time series for the two pitch classes in the two sessions, pooling pitch levels 1 - 3 and 4 - 6 respectively independent of velocity and density values. This trial division corresponded to the trained category boundary (see **Figure 2.1B**). The hemodynamic response delay was corrected for by convolving the predicted time courses with a canonical (double gamma) hemodynamic response function (HRF). We performed both single-subject and group (fixed-effects) analyses of the contrast 'high pitch' vs. 'low pitch' both for the pre- and post-learning session. Thresholds for contrast maps were corrected for multiple comparisons based on False Discovery Rate ($q = 0.05$).

Multivariate data analysis. All multivariate pattern analyses were performed on a single-subject basis. Activity patterns were estimated trial by trial (72 x 3) in an anatomically defined auditory cortex mask, covering the superior temporal gyrus (STG) including Heschl's gyrus (HG) and its adjacency (i.e. its anterior and posterior borders reaching into planum polare (PP) and planum temporale (PT)) as well as the superior temporal sulcus (STS). Anatomical masks were delineated on an inflated cortex mesh for each subject and hemisphere separately to account for differences in gross anatomy. At each voxel, the trial response was extracted by fitting a GLM with one predictor for the expected BOLD response and one predictor accounting for the trial mean. A multi-voxel pattern was defined from the response-related beta coefficients (De Martino et al., 2008; Formisano et al., 2008). The shape of the hemodynamic response function was optimized per subject.

The multi-voxel response patterns to the different sound classes were analyzed by means of linear Support Vector Machines (SVM) in combination with an iterative voxel selection algorithm (RFE, De Martino et al., 2008) to derive the most informative voxels. We followed two different strategies to label each single trial response pattern. In a first approach, trials were divided based on the trained dimension: Trials with pitch levels 1 - 3 and 4 - 6 were assigned to class 1 and class 2 respectively, independent of the other stimulus dimensions. In an alternative control approach, trials were labeled according to the untrained dimension (i.e. velocity), resulting in two classes comprising trials with either slow (1 - 3 cycles/second) or fast (4 - 6 cycles/second) velocity values, irrespective of pitch and density. Both strategies resulted in 36 trials per class in each run. In four of the eight subjects - to estimate the beta-coefficients in an appropriate time window of four TRs per trial - we needed to remove the last trial of each run due to insufficient data supply. A trial from the respective other class was equally deleted to balance the number of trials per class resulting in 35 trials.

For classifier training, trials were divided into a training and a test set using a leave-one run-out approach resulting in three different splits. Two runs (i.e. 72 trials per class) were used for classifier training while the remaining run (i.e. 36 trials per class) was used to assess classifier performance and test its generalization ability. This procedure was repeated for the number of splits. This validation procedure avoids potential overfitting of the model to irrelevant fluctuations in the training data. Final accuracy values at each voxel selection level were computed as the mean over the three splits for the test data set only. Each split of the leave-one run-out cross validation procedure included a univariate (GLM-based) feature selection based on the training data only. Using one predictor per class, we selected the 5000 most active voxels (overall main effect, *F*). This constrains the classification procedure to those voxels that exhibit a general response to the employed stimuli and limits the classification to an equal number of voxels in each subject (for details see De Martino et al., 2008). This was followed by 160 iterations of the RFE algorithm. In each of the iterations, a different subset of the training trials (95 %) was used to train the classifier

and to retrieve voxels' discriminative weights. These weights provide information about the relative contribution of voxels to class discrimination. Classification accuracy at each level was assessed on the independent test data set. After four consecutive trainings, the ranked discrimination weights were averaged and the lowest 10 % were discarded while the rest was used to retrain the classifier. This procedure resulted in 40 voxel selection levels per split.

To assess whether our classification accuracies significantly differed from chance level we employed a permutation test (Nichols and Holmes, 2002). For this purpose, the same RFE procedure used for the experimental protocols was repeated 100 times per subject, session, and trial division (i.e. trained/untrained), with scrambled trial labels (using the randperm function in Matlab). Classification accuracies for permutations are based on the maximum accuracy across 40 RFE levels (averaged across splits) in each permutation averaged over 100 iterations for each subject and fMRI session separately. This procedure controls for the potential bias in the accuracy estimation introduced by considering the best feature selection level.

To investigate the cortical regions involved in discrimination of the newly learned categories, group discriminative maps were visualized on an average cortex reconstruction following cortex-based alignment of single-subject discrimination maps. In **Figure 2.3B**, we display those voxels, which consistently survived at least 10 of the 40 RFE selection levels in six out of eight subjects. Maps were corrected by applying a cluster size threshold of 25 mm^2. An identical procedure for the fMRI data collected before learning did not lead to consistent voxels.

Learning-induced fMRI pattern changes and relation to behavior. To examine the relation between learning-induced changes in fMRI patterns and behavioral changes we performed the following analysis. First - for each subject and for both pre- and post- fMRI sessions - we defined a prototypical response pattern for category A and B by considering the average response pattern (training data) for pitch levels 1 - 3 and 4 - 6 respectively in the 500 voxels with the largest SVM-weights in the 10th voxel selection level.

Second, we correlated the prototypical response patterns with the response patterns for each individual pitch level (1 - 6), estimated from the same voxels and using test trials only. Per subject, thus we obtained four vectors describing the similarity of the response patterns to the prototypical response to category A and B, before and after learning, i.e. values $c_i(pA^{pre})$, $c_i(pB^{pre})$, $c_i(pA^{post})$, $c_i(pB^{post})$, where i = 1..6 indicates the pitch level. To remove the intrinsic correlation between responses, difference scores were calculated in each subject as $d_i^{pre} = c_i(pB^{pre}) - c_i(pA^{pre})$, $d_i^{post} = c_i(pB^{post}) - c_i(pA^{post})$ after all correlation values were transformed using Fisher z. The curve plotted in **Figure 2.5** indicates the differences in fMRI pattern similarities between pre- and post- fMRI session, obtained by fitting the difference $d_i^{post} - d_i^{pre}$ (by eq. 1), averaged across subjects. Analogously, we computed the post – pre difference in behavioral identification functions (% B responses) to reveal the learning-induced changes in perceptual similarity. For visualization purposes both fMRI and behavioral curves were standardized using the z-transformation.

3 | RESULTS

Behavioral results. Average categorization performance reflected successful learning of pitch classes in two training days (corresponding on average to 324 feedback trials). Accuracy - as measured in non-feedback test blocks prior to training and after one and two/three training blocks - increased gradually and significantly ($F_{2,14} = 31.10$, $p < .001$) with training (**Figure 2.2A**). **Figure 2.2B** shows that prior to learning, the average sound identification curve was rather flat and had a small amplitude (estimated parameters of the fit: $a = 33.51$; $b = 40.6$; $c = 0.46$; $d = 4.22$) reflecting the ambiguity of the classes with respect to the sound dimensions and the continuous nature of ripple pitch. With learning, the curve expanded along the y-axis, indicating that the category extremes were classified with higher confidence, and changed into a steep sigmoid shape with a sharp transition at the category boundary ($a = 98.3$; $b = 0$; $c = 0.44$; $d = 3.41$), a characteristic signature of CP (Harnad, 1987). Average goodness of fit

expressed in adjusted R^2 was .96 and .99 for pre and post learning respectively.

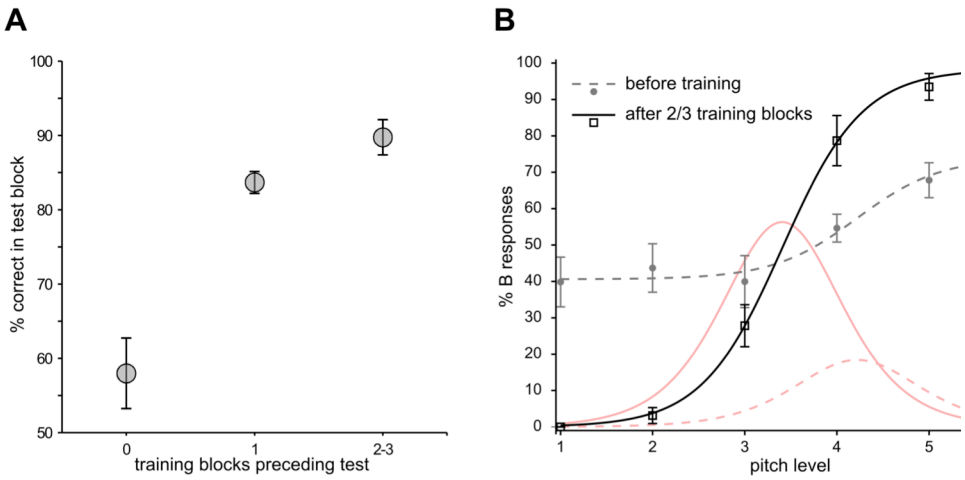
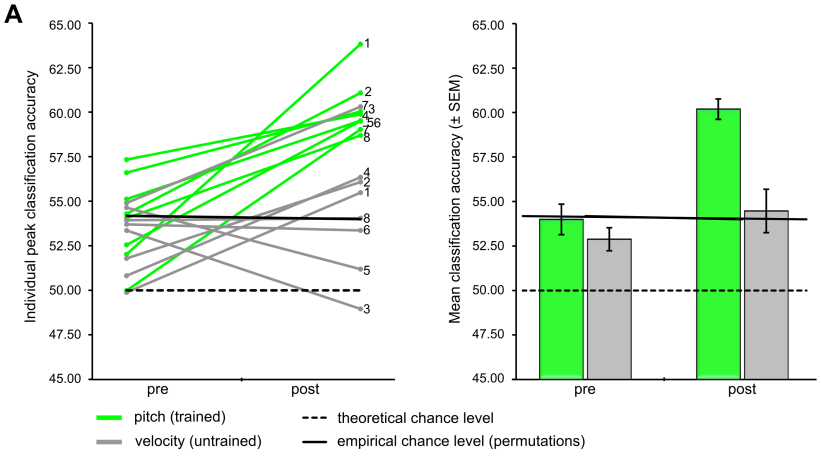


Figure 2.2 Group behavioral results (data are represented as mean +/- *SEM*). A, Categorization accuracy in three non-feedback test blocks before training, and after 1 and 2/3 training blocks respectively. **B**, Identification functions (curve-fitting results and original data points before training and after 2/3 training blocks). A logistic function (eq. 1) was fitted to the mean probabilities to categorize a sound as 'B' along the pitch continuum. The derivative of the respective curves is indicated in light red to highlight the shift and steepening of the category boundary reflected by the maximum of the function.

Imaging results - univariate statistical analysis of fMRI data. Ripple sounds significantly activated extended regions on bilateral superior temporal cortex. FMRI responses included large parts of the STG, including HG, Heschl's sulcus (HS) and PT as well as smaller portions of the STS and insular cortex. Univariate contrasts between trained categories did not yield any significant response differences for the group (fixed-effect) and for each single subject separately (FDR-corrected threshold, q = .05) neither before nor after learning. These results are consistent with the hypothesis that learning may induce subtle neural changes without significant changes in overall activation (Ohl et al., 2001; Schnupp et al., 2006).


A
Individual peak classification accuracy
65.00
62.50
60.00
57.50
55.00
52.50
50.00
47.50
45.00
pre
post
Mean classification accuracy (± SEM)
pitch (trained)
velocity (untrained)
theoretical chance level
empirical chance level (permutations)

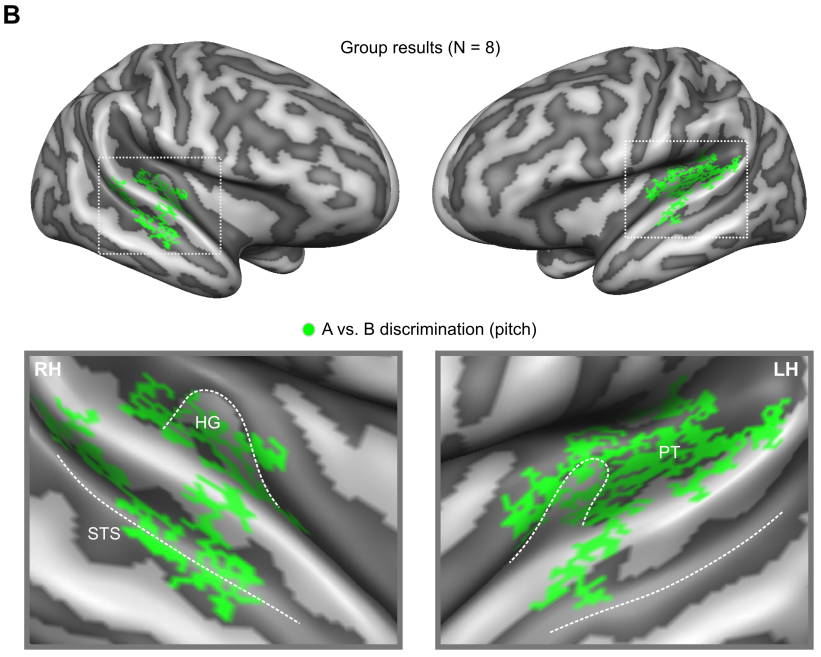

B
Group results (N = 8)
A vs. B discrimination (pitch)
RH
HG
STS
LH
PT

Figure 2.3 Imaging results – fMRI pattern analysis. A, The left panel shows individual peak classification accuracies based on fMRI data prior to category training and after successful category learning for the two types of stimulus space divisions (trained vs. untrained) and the respective trial labeling. Individual subjects are indicated with numbers on the right of each line. Group-averaged classification accuracies are shown on the right. Mean accuracies are computed from the individual peak values across voxel elimination levels for each session and trial division. Theoretical and average empirical chance levels (estimated through permutation) are displayed as dashed black and solid black lines respectively. **B**, Group discrimination maps based on the post-learning fMRI data for the trained stimulus division (i.e. 'low pitch' vs. 'high pitch'), displayed on an average reconstructed cortical surface after cortex-based realignment. A certain cortical location (vertex) was color-coded when it survived at least ten of the feature elimination levels in at least 6 out of 8 subjects (cluster size threshold 25 mm^3).

Imaging results - decoding of novel sound categories from fMRI patterns. We compared pre- and post-training classifier performance on unlabeled trials after the algorithm had been trained with a subset of trials labeled either according to the trained (pitch) or untrained (velocity) sound dimension, irrespective of the other sound features. We thereby assessed the correspondence of the fMRI pattern discrimination with the behavioral learning rule. A repeated measures ANOVA revealed a significant interaction between fMRI session and trial labels ($F_{2,14} = 11.82$; $p = .001$, **Figure 2.3A**). Before category learning, the classifier did not succeed in distinguishing two sound classes based on either dimension. Classification accuracy for test trials did not significantly differ from empirical chance level, estimated with permutation. After subjects were trained, average classification accuracy across 8 subjects reached 60.19 % for the trained sound classes (pitch) and only 54.47 % for the untrained sound classes (velocity). Two a priori hypotheses were tested with Bonferroni corrected alpha levels of .025. The pairwise comparison of pitch classification accuracies before and after training revealed a significant increase in accuracy with category learning ($t_7 = 5.67$, $p = .001$). In the post-training session, accuracies for pitch discrimination were significantly above the empirical chance level of 54 % ($t_7 = 9.58$, $p < .001$). In seven out of eight subjects, the classification accuracy for trained pitch classes significantly (p

≤ .05) differed from accuracies obtained with permuted trial labels (**Figure 2.4**).

Importantly, category training affected perceptual similarity expressed in sound identification curves and fMRI pattern similarity derived from correlation measures in an analogues manner. After category learning, neural response patterns for sounds with higher pitch (pitch levels 4, 5, 6) correlated with the prototypical response pattern for class B more strongly than class A, independent of other acoustic features. The profile of these correlations on the pitch continuum closely reflected the sigmoid shape of the behavioral category identification function (**Figure 2.5**). On average, these learning-induced pattern changes strongly correlated with the changes in behavioral sound categorization (r = .91, p = .01).

Imaging results – group discrimination maps. Voxel patterns discriminative for the learned pitch classes were distributed bilaterally over the auditory cortex and included regions of the primary and early auditory areas (on HG and adjacent regions). Both hemispheres revealed activation clusters in the posterior lateral portion of HG (corresponding approximately to MNI (Montreal Neurological Institute) coordinates ±45, -20, 12) extending beyond its posterior border into HS and PT (mainly left hemisphere, -45, -30, 12) and anteriorly into the first transverse sulcus (FTS) (see **Figure 2.3B**). Especially in the right hemisphere, additional clusters were found on anterior lateral HG (48, -13, 4) and extended portions of the middle STG/STS (45, -19, -5). These voxels were highly consistent across subjects (6 out of 8) and stable over at least 10 elimination levels.

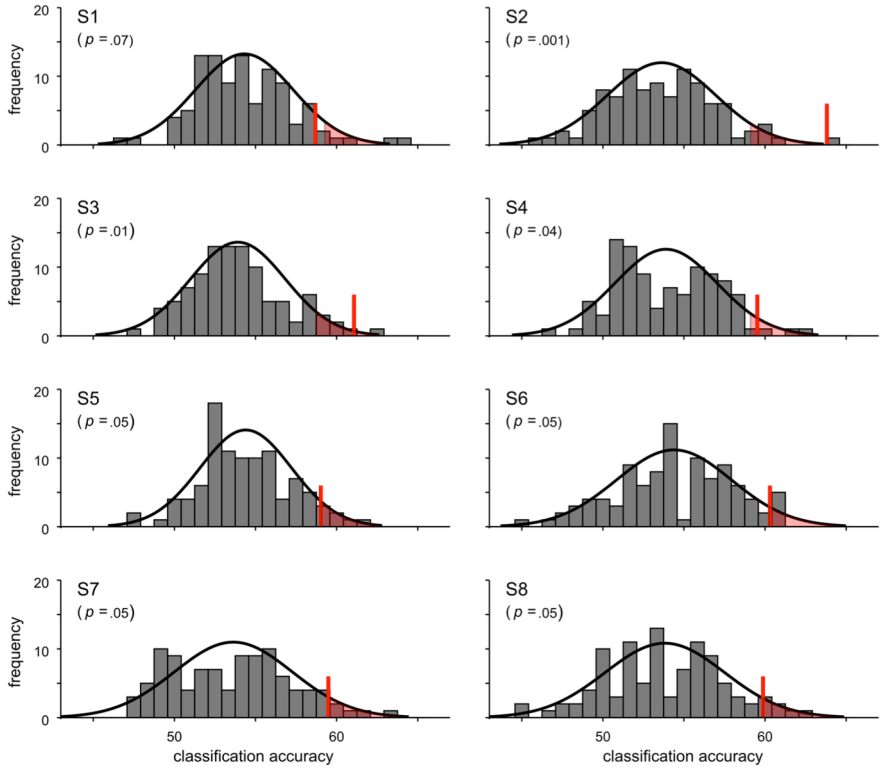


Figure 2.4 Distribution of classification accuracies obtained with permuted trial labels. The values reflect the maximum classification accuracy across 40 RFE levels (averaged over splits) for 100 permutations for each subject (N = 8) separately. The normal curve is defined by the mean and standard deviation of the underlying distribution. The red shading reflects the 95 % confidence interval. The red marker indicates the actual accuracy obtained with trial labels according to the trained (i.e. pitch) dimension. P-values (extracted from the cumulative distribution function) reflect above-chance ($p \leq .05$) classification in 7 of 8 subjects.

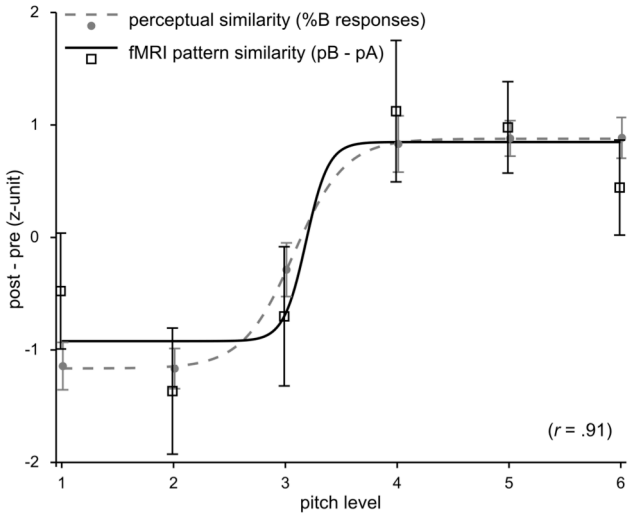


Figure 2.5 Changes in pattern similarity and behavioral identification curves. The learning-induced change in fMRI pattern similarity along the pitch continuum (levels 1 - 6) is illustrated by correlation difference scores (d) contrasted between post- and pre-learning sessions (please refer to the Method section for details). Behavioral data analogously correspond to the post - pre difference in identification functions (% B responses). Data are visualized in z-units and represent the group mean +/- *SEM*. Lines reflect the fit with the sigmoid function (eq. 1) used for behavioral data analysis (see *Curve fitting procedure*). Markers are displayed with a slight offset to increase visibility. Pearson's correlation coefficient (*r*) indicates strong correspondence between behavioral and neural measures.

4 | DISCUSSION

In this fMRI study, we used multi-voxel pattern analysis to reveal changes in sound representations induced by the formation of new perceptual categories in human auditory cortex. We trained subjects to dissect a multidimensional sound space based on one relevant feature and measured neural responses to the passive exposure to a sound continuum before and after successful category learning.

Listeners successfully learned the new sound categories as reflected in their categorization accuracy and the shape of the category identification function. The gradual increase of categorization performance across training blocks suggests that a sudden insight into the relevant acoustic dimension

alone was insufficient to achieve precise categorization. Instead, perceptual learning (Ahissar, 1999) at the category boundary was required for optimal classification. In accordance with previous studies (Smits et al., 2006; Goudbeek et al., 2009) categorization performance transferred well from the Gaussian training distributions to the continuous stimulus space and persisted despite lack of feedback. This demonstrates the generalization of the learned categories to novel sounds without distributional cues indicative of the category structure or direct reinforcement in the form of corrective feedback. This abstraction process is considered fundamental to categorization (Kéri, 2003). The sigmoid shape of the category identification function after training resembled the labeling data from natural phoneme categories (Liberman et al., 1957). The formation of the category boundary separating the two pitch classes required the abstraction of discrete classes from continuous pitch information and the mapping of pitches to different labels on either side of the classification boundary. Moreover, subjects had to ignore the additional irrelevant spectral and temporal variations and select only pitch as the basis for the development of abstract representations of pitch classes. Perceptual invariance of sets of objects classified as belonging to the same category (despite variations in some aspects) is considered a vital function underlying object recognition (Jüttner & Rentschler, 2008; Walker et al., 2011).

Crucially, prior to learning, the abstract pitch categories could not be decoded from the auditory cortex. This argues against pre-existing representations of our sound categories and suggests that feature mapping alone is insufficient for categorical representations. Frequencies discriminable in tonotopic maps usually lie much further apart and reflect voxel's relative preference resulting from best-frequency analysis (i.e. color coding of frequency at which the response is maximum) rather than significant frequency contrasts (Formisano et al., 2003). Furthermore, the pitch classes contrasted in our analysis are characterized by large within-class variability, not only in the irrelevant dimensions (velocity and density) but also along the relevant dimension (3 pitch values are grouped into one class). After learning, the classifier correctly assigned activation patterns in

the auditory cortex to their corresponding pitch class, independent of the other spectro-temporal variations present in the sounds. These results suggest the development of discriminative response patterns for the pitch classes with learning. It should be noted that category learning did not affect the representation of all sound features but selectively enhanced the differences in the behaviorally relevant dimension at the learned category boundary. This important differentiation therefore excludes repeated stimulus exposure as a potential cause of increased classifier performance (Seitz and Watanabe, 2003) and provides direct evidence for specific representational changes in human auditory cortex with category learning.

The widespread activation of auditory areas can be attributed to the complex spectro-temporal structure of the employed rippled sounds, which engage a multitude of functional processing areas (Langers et al., 2003; Schönwiesner and Zatorre, 2009). Given the identical stimulus sets for pre- and post-learning fMRI sessions and the uniform distribution of the employed test sounds, the changes in sound representations essentially rely on perceptual re-interpretations of the same acoustic input induced by category learning. Our results demonstrate the flexibility of sound representations in early auditory areas and the ability of auditory cortical neural populations to adapt relatively quickly to situation-dependent changes in the environment. This further supports the notion that these areas serve higher-order sound analysis beyond feature-extraction in line with previous reports (Nelken, 2004). The resemblance of the activity pattern similarity and the perceptual sound similarity as reflected in the identification curves obtained outside the scanner suggests a link between perception and neural population coding. The good fit to the specified sigmoid function (adjusted R^2 of .78 for the fMRI and .99 for the behavioral data), generally used to model categorical data, suggests that continuous pitch information is represented categorically in distributed multi-voxel patterns after learning.

Discriminative maps resulting from multivariate analyses should be considered as a whole rather than localized hotspots. The essence of pattern analysis using linear classifiers is the weighted contribution of multiple

voxels rather than the specialization of a particular cortical region. Yet, relating the most consistently informative locations with previous fMRI reports is useful to integrate our data in current knowledge.

The lateral posterior part of HG and the posteriorly adjacent areas have previously been shown to code perceptual states rather than purely acoustic differences of sounds (Kilian-Hütten et al., 2011). Furthermore, these areas have been used to reliably decode speaker information from natural and variable speech sounds (Formisano et al., 2008). Thus, they seem to play an important role in abstract and goal-directed representation of sounds. Activation in the right STS/STG is strongly related to vocal processing (Belin et al., 2000; Belin & Zatorre, 2003; Formisano et al., 2008), specifically the extraction of speaker identity and other paralinguistic information. As our sounds were non-harmonic complexes, the similarity to vocal sounds is rather small, however, voice identification is predominantly based on the extraction of the fundamental frequency (Belin et al., 2004; Baumann and Belin, 2010), which is the underlying acoustic dimension upon which ripple classification was based in our experiment. The right anterior lateral HG has been described to be involved in pitch analysis (Warren & Griffiths, 2003; Barrett & Hall, 2006). The recruitment of areas specialized in pitch processing is in line with the previously proposed concept of reallocation of resources according to task demands (Brechmann and Scheich, 2005). Altered representations of identical visual stimuli depending on the task-relevant features (Mirabella et al., 2007) and increased selectivity for diagnostic features (Sigala & Logothetis, 2002; De Baene et al., 2008) have previously been demonstrated in monkeys during active categorization. Despite the lack of control over the subjects' performance during scanning, none of our subjects reported to have actively categorized the sounds. The finding of learning-induced modifications of stimulus representations in our study during passive listening suggests that task-related processes shape stimulus representations beyond the scope of the learning environment, yielding a multi-purpose enhancement of neural sensitivity for the relevant stimulus differences. This provides neurophysiological support for the effects of "acquired distinctiveness/equivalence", where relevant stimulus

dimensions attain elevated discriminability while perceptual sensitivity for irrelevant dimensions is decreased after category learning (Goldstone, 1994). The emphasis of category-relevant processes at the expense of category-irrelevant processes at the level of the auditory cortex may increase overall efficiency and facilitate read-out in higher order regions, conforming with theories of sparse coding (Olshausen & Field, 2004).

Contrary to predictions from earlier reports (Desai et al., 2008; Leech et al., 2009), increased categorical processing of ripple sounds did not engage left posterior STS. This argues against a generic role of these speech-related areas in categorical processing but rather proposes that categorically perceived sounds specifically recruit left STG/STS for mapping onto highly abstract and overlearned phonemic representations if they share spectro-temporal speech characteristics.

Despite the prevalent view that the PFC is the main site of category representations, in the visual domain the contribution of frontal and higher occipito-temporal and parietal areas in category learning remains under debate (Kourtzi & Connor, 2011). While comparisons between the auditory and visual domain might be limited by general cortical processing differences, our results provide direct evidence for representations of abstract sound categories already at early levels of the auditory processing hierarchy. While the current experiment cannot exclude the contribution of the PFC in categorical sound processing, recent evidence in humans suggests that the PFC is predominantly involved in rule learning and specifically recruited in the context of an active categorization task (Boettiger & D'Esposito, 2005; Li et al., 2009). The passive design employed in the current study seems particularly suitable to reveal learning-dependent changes in the representations of sound categories in early processing areas rather than decision-related processes in the PFC.

To conclude, our data present direct evidence in humans for learning-induced formation of categorical sound representations in early auditory areas. While responses to a psychophysical sound continuum could not be distinguished prior to learning, few days of category training sufficed to reliably decode newly formed pitch categories from distributed response

patterns in pitch-encoding areas in absence of an active categorization task. Our results are consistent with animal studies and demonstrate that fMRI pattern analyses are eligible to reveal subtle changes in sound representations otherwise inscrutable to conventional contrast-based methods. Furthermore, our findings provide an important demonstration of the plastic nature of sound representations at early processing stages in human auditory cortex.

REFERENCES

Ahissar M (1999) Perceptual learning. Curr Dir Psychol Sci 8: 124–128.

De Baene W, Ons B, Wagemans J, Vogels R (2008) Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. Lern Mem 15: 717–727.

Barrett DJK, Hall DA (2006) Response preferences for "what" and "where" in human non-primary auditory cortex. NeuroImage 32: 968–977.

Baumann O, Belin P (2010) Perceptual scaling of voice identity: common dimensions for different vowels and speakers. Psychol Res 74: 110–120.

Belin P, Fecteau S, Bédard C (2004) Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 8: 129–135.

Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14: 2105–2109.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403: 309–312.

Boettiger CA, D'Esposito M (2005) Frontal networks for learning and executing arbitrary stimulus-response associations. J Neurosci 25: 2723–2732.

Brechmann A, Scheich H (2005) Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cereb Cortex 15: 578–587.

Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J Neurophysiol 85: 1220–1234.

Desai R, Liebenthal E, Waldron E, Binder JR (2008) Left posterior temporal regions are sensitive to auditory categorization. J Cogn Neurosci 20: 1174–1188.

Formisano E, Kim DS, Di Salle F, Van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40: 859–869.

Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322: 970–973.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291: 312–316.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2003) A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J Neurosci 23: 5235–5246.

Glasberg BR, Moore BC (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47: 103–138.

Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum Brain Mapp 27: 392–401.

Goldstone R (1994) Influences of categorization on perceptual discrimination. J Exp Psychol Gen 123: 178–200.

Goudbeek M, Swingley D, Smits R (2009) Supervised and unsupervised learning of multidimensional acoustic categories. J Exp Psychol Hum Percept Perform 35: 1913–1933.

Guenther FH, Husain FT, Cohen MA, Shinn-Cunningham BG (1999) Effects of categorization and discrimination training on auditory perceptual space. J Acoust Soc Am 106: 2900–2912.

Harnad S (1987) Categorical perception: The groundwork of cognition (Harnad S, ed). Cambridge, UK: Cambridge UP.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293: 2425–2430.

Jüttner M, Rentschler I (2008) Category learning induces position invariance of pattern recognition across the visual field. Proc Biol Sci 275: 403–410.

Kéri S (2003) The cognitive neuroscience of category learning. Brain Res Rev 43: 85–109.

Kilian-Hütten N, Valente G, Vroomen J, Formisano E (2011) Auditory cortex encodes the perceptual interpretation of ambiguous sound. J Neurosci 31: 1715–1720.

Kourtzi Z, Connor CE (2011) Neural representations for object perception: structure, category, and adaptive coding. Annu Rev Neurosci 34: 45–67.

Kowalski N, Depireux DA, Shamma SA (1996a) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J Neurophysiol 76: 3503–3523.

Kowalski N, Depireux DA, Shamma SA (1996b) Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. J Neurophysiol 76: 3524–3534.

Langers DRM, Backes WH, Van Dijk P (2003) Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples. NeuroImage 20: 265–275.

Leech R, Holt LL, Devlin JT, Dick F (2009) Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. J Neurosci 29: 5234–5239.

Levitt H (1971) Transformed up-down methods in psychoacoustics Tobias J V, ed. J Acoust Soc Am 49: 467–477.

Li S, Mayhew SD, Kourtzi Z (2009) Learning shapes the representation of behavioral choice in the human brain. Neuron 62: 441–452.

Li S, Ostwald D, Giese M, Kourtzi Z (2007) Flexible coding for categorical decisions in the human brain. J Neurosci 27: 12321–12330.

Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol 54: 358–368.

Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR (2010) Specialization along the left superior temporal sulcus for auditory categorization. Cereb Cortex 20: 2958–2970.

De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. NeuroImage 43: 44–58.

McMurray B, Spivey M (2000) The categorical perception of consonants: the interaction of learning and processing. Proceedings of the Chicago Linguistics Society 34: 205–220.

Mirabella G, Bertini G, Samengo I, Kilavik BE, Frilli D, Della Libera C, Chelazzi L (2007) Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. Neuron 54: 303–318.

Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14: 474–480.

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15: 1–25.

Ohl FW, Scheich H, Freeman WJ (2001) Change in pattern of ongoing cortical activity with auditory category learning. Nature 412: 733–736.

Olshausen BA, Field DJ (2004) Sparse coding of sensory inputs. Curr Opin Neurobiol 14: 481–487.

Schnupp JWH, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J Neurosci 26: 4785–4795.

Schönwiesner M, Zatorre RJ (2009) Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proc Natl Acad Sci U S A 106: 14611–14616.

Seger CA, Miller EK (2010) Category learning in the brain. Annu Rev Neurosci 33: 203–219.

Seitz AR, Watanabe T (2003) Is subliminal learning really passive? Nature 422: 2003.

Selezneva E, Scheich H, Brosch M (2006) Dual time scales for categorical decision making in auditory cortex. Curr Biol 16: 2428–2433.

Shamma S (1996) Auditory cortical representation of complex acoustic spectra as inferred from the ripple analysis method. Netw Comput Neural Syst 7: 439–476.

Sigala N, Logothetis NK (2002) Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415: 318–320.

Smits R, Sereno J, Jongman A (2006) Categorization of sounds. . J Exp Psychol Hum Percept Perform 32: 733–754.

Staeren N, Renvall H, De Martino F, Goebel R, Formisano E (2009) Sound categories are represented as distributed patterns in the human auditory cortex. Curr Biol 19: 498–502.

Walker KMM, Bizley JK, King AJ, Schnupp JWH (2011) Cortical encoding of pitch: recent results and open questions. Hear Res 271: 74–87.

Warren JD, Griffiths TD (2003) Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J Neurosci 23: 5799–5804.

Wetherill GB, Levitt H (1965) Sequential estimation of points on a psychometric function. Br J Math Stat Psychol 18: 1–10.

Chapter 3

AUDIO-TACTILE AND AUDIO-VISUAL LEARNING PROMOTE CROSSMODAL REORGANIZATION AT DIFFERENT NEURAL PROCESSING LEVELS

Ley, A., Kaas, A., Formisano, E., & Vroomen, J. (in preparation). Audio-tactile and audio-visual learning promote crossmodal reorganization at different neural processing levels.

Sie mag Musik nur, wenn sie laut ist
Wenn der Boden unter den Füßen bebt
Dann vergisst sie, dass sie taub ist

~ Herbert Grönemeyer

ABSTRACT

While there is a considerable amount of evidence for crossmodal plasticity during development, research on the effect of short-term multisensory experience on neural sound representations in the mature human brain is scarce. This study combined two matched multisensory training paradigms with functional magnetic resonance imaging (fMRI) to investigate learning-induced plasticity of neural representations of sounds. We trained participants to associate a large number of artificial sounds with one visual (group 1) or vibrotactile (group 2) stimulus appearing at two distinct spatial locations and compared group-specific differences in pre- and post-learning sound activations. Our results show that after audio-visual training, sounds alone activated a fronto-parietal processing network and engaged parietal areas, which were unresponsive to the same sounds prior to learning. In contrast, effects of audio-tactile learning were restricted to the temporal lobe with increasing sound responses in the auditory belt and posteriorly adjacent regions. This discrepancy suggests that audio-visual and audio-tactile interaction during learning occurs at different processing levels resulting in discriminable multisensory representations that are activated by the same sounds after training. The findings are in line with mounting evidence for low-level convergence of auditory and somatosensory stimuli in the auditory cortex and suggest that crossmodal plasticity is not restricted to ecologically valid associations but occurs for newly learned arbitrary bimodal stimulus pairs.

1 | INTRODUCTION

Learning to interpret and categorize sounds causes rapid reorganization of their neural representation (for review, see Scheich et al., 2011). Recent studies in animals (Ohl, Scheich, & Freeman, 2001) and humans (Ley et al., 2012) have shown that activation patterns in early auditory cortex adapt to the context and task, forming abstract representations of the behaviorally relevant sound information. However, so far, learning-induced changes in sound representations have only been investigated in the context of unisensory training paradigms, neglecting the naturally occurring crossmodal influence and the superior efficacy of bimodal learning paradigms (Seitz, Kim, & Shams, 2006; Shams & Seitz, 2008). Integrating the contextual information from vision or touch provides vital cues for sound identification. Especially in noisy environments, human observers naturally and effortlessly make use of this qualitatively distinct but complementary information (Schürmann, Caetano, Jousmäki, & Hari, 2004; Sumby & Pollack, 1954). Crucially, multisensory exposure can exhibit long-lasting effects on later processing of the constituent unimodal parts. In fact, unimodal stimuli with multisensory pasts can evoke enhanced activation and connectivity of the sensory-specific processing regions (Murray, Foxe, & Wylie, 2005; von Kriegstein & Giraud, 2006) and lead to crossmodal recruitment of neural assemblies specialized for processing the associated but absent stimulus component (Calvert et al., 1997; Pekkola et al., 2005; von Kriegstein et al., 2005). These findings suggest that the cortex is prone to experience-dependent crossmodal plasticity. Yet, while these effects have been demonstrated for ecologically valid associations such as voices and faces, it remains uncertain whether these results would generalize to newly learned multisensory associations between arbitrary stimulus pairs. Critically, what we know about learning-dependent crossmodal plasticity is largely based on audio-visual interactions, while little attention has been paid to other crossmodal combinations. Accumulating evidence points towards fundamental differences in the mechanisms underlying audio-visual and audio-tactile integration (Schroeder & Foxe, 2002). Due to the lack of

studies that compared different multisensory learning paradigms with novel stimuli, it remains unknown to which degree the observed crossmodal effects are modality-specific.

This fMRI study directly compared the effects of audio-visual and audio-tactile association training on the representations of the same auditory stimuli, controlling for perceptual experience, training duration, as well as task complexity. We employed a complex many-to-one crossmodal association task in which participants learned to match a large number of initially meaningless sounds (ripples) to one visual (group 1) or vibrotactile (group 2) stimulus occurring at one of two distinct spatial locations. To investigate learning-induced changes in unisensory responses, participants were scanned while passively listening to the sounds in isolation prior to and following three days of association training. We hypothesized that audio-visual and audio-tactile association learning leads to strengthened connectivity between the feature representations in the respective modality, which in turn should result - post-learning - in enhanced sound activation in visual and tactile processing regions, respectively.

2 | MATERIAL AND METHODS

Participants. Fourteen participants recruited from Maastricht University, the Netherlands, volunteered for the fMRI experiment and were reimbursed for their participation. All participants gave written informed consent before participation in the study. They had no history of hearing-loss or neurological disorders, and normal or corrected-to-normal vision and no hand injuries. Seven participants (5 females; mean age: 28 years) were randomly allocated to the audio-visual (AV) training group; the other seven participants (5 females; mean age: 25 years) were allocated to the audio-tactile (AT) training group. The Ethical Committee Psychology at Maastricht University granted approval for the study.

Auditory stimuli. We used constant moving ripples (**Figure 3.1A**) (Kowalski et al., 1996a) to create novel sound categories, generated with the Matlab

(MathWorks) NSL toolbox, available at http://www.isr.umd.edu/Labs/NSL/Software.htm. The sounds differed in three spectro-temporal dimensions corresponding to the fundamental frequency of the complex (f_0 in Hz), the spectral modulation density (Ω in cycles/octave, cyc/oct) and the temporal modulation rate (ω, in Hz). The spacing between frequency components of the ripples was 1/16 octave (bandwidth of ripple complex: 4 octaves); f_0 values varied between ~150-250 Hz and were determined by the category distributions (see **Figure 3.1D**, description follows below). Ripple density was either 1 or 2 cyc/oct with a modulation depth of 1. The phase of the spectral envelope was shifted downwards (start phase of 0) with constant speed (ω: 1,2,3,4,5 or 6 Hz) over the entire duration of the sound (1 s). These low modulation rates reflect naturally occurring modulation such as in speech (Chi et al., 1999). Ripples were down-sampled to 44100 Hz using Audacity (Free Software Foundation, Inc., Boston, USA), ramped (15 ms linear amplitude slope at on- and offset), and normalized according to their root mean square values (rms). Sounds were delivered binaurally through MR-compatible in-ear headphones (Sensimetrics).

Visual stimuli. Visual stimuli consisted of circular patches (diameter 2° visual angle) with oblique (45°) black-and-white square-wave gratings (spatial frequency 4 cycles/°) with a constant phase shift of 1/4 cycle (90°). The gratings were presented on a uniform grey background at two different polar angles in the upper-left visual field (at 110° and 160°) and at 5° eccentricity from the central fixation cross (size: 1°, see **Figure 3.1B**). Those stimulus locations were chosen to match the left lateralized tactile stimulation without creating a clear 'high-low'-dissociation.

Tactile stimuli. Vibrotactile stimuli were delivered to the tip segment of the little (T1) and index (T2) finger of the left hand via a piezoelectric mechanical stimulation device (Piezostimulator, QuaeroSys Medical Devices, Schotten, Germany). Each finger was positioned on a stimulation module (**Figure 3.1C**) consisting of a 2 x 5 pin matrix (pin diameter: 1 mm,

center distance between pins: 2.5 mm). All ten pins of one module were triggered simultaneously to a maximum height of 0.9 mm with a frequency of ~33 Hz (square wave, timing accuracy 0.5 ms) for the duration of one second, matching the sound duration. The stimulation modules were fixed at the fingers with sport bandages to ensure stable pressure throughout the experiment. Participants rested their hands in a comfortable position on their body while lying in the scanner. The combination of in-ear headphones with foam tips used for sound delivery and over-ear headphones used for communication masked the soft buzzing sound produced by the pin movement of the piezostimulator.

Category distributions. New sound categories were defined by ripple pitch (determined by f_0) irrespective of the other sound parameters (Ω and ω). Ambiguous sounds close to the category boundary (f_0 = 200 Hz) were omitted during category training. Pitch values were linearly sampled from two Gaussian distributions in logarithmic frequency space, with category means (μ_A and μ_B) in a distance of 1/4 octave from the category boundary (see red and blue dots in **Figure 3.1D**). The standard deviation (σ) of the curves was constrained by 1/8 octave. Each sound category comprised 72 sounds with six different pitch values (clustered around μ_A = 168.18 Hz and μ_B = 237.84 Hz), six different temporal modulation rates and two different modulation densities.

To examine the level of abstraction of two distinct pitch categories from the continuous variation of the underlying physical feature (f_0), we created a different set of test sounds (see black crosses in **Figure 3.1D**) with six f_0 values sampled in equidistant steps (in log space) between category means (μ_A and μ_B) crossing the category boundary. Each of these six test sounds was presented with six temporal modulation rates and two modulation densities. These test sounds were never combined with corrective visual feedback and were used to test crossmodal matching performance at two different learning phases (day1 and day3) and to map sound representations in the main fMRI experiment.

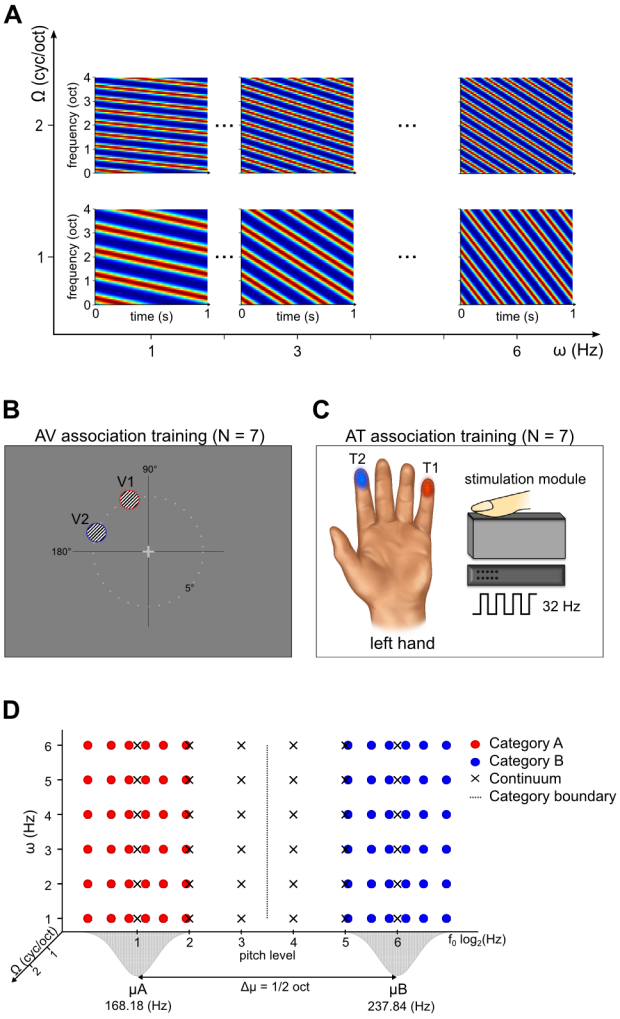

A
Ω (cyc/oct)
frequency (oct)
time (s)
ω (Hz)
B
AV association training (N = 7)
V1
V2
90°
180°
5°
C
AT association training (N = 7)
T2
T1
stimulation module
32 Hz
left hand
D
Category A
Category B
Continuum
Category boundary
ω (Hz)
Ω (cyc/oct)
pitch level
f0 log2(Hz)
μA
168.18 (Hz)
Δμ = 1/2 oct
μB
237.84 (Hz)

Figure 3.1 Stimulus design. A, Schematic illustrations of the ripple sounds. Example spectrograms represent the two spectral modulation densities (y-axis) and the variation of the spectral modulation rate (x-axis). **B**, Depiction of the visual gratings used for AV training. Axes and circle outlines were not shown during the experiment. Low-pitch sounds (shown in red in D) were associated with a polar angle of 110°; high-pitch sounds were associated with the 160° stimulus location. **C**, Vibrotactile stimulation locations on the little and index finger of the left hand as well as schematic representation of a stimulation module with 2 x 5 pin matrix. Red and blue colors reflect the learned association with the corresponding sounds shown in D; i.e. low-pitch sounds were correctly paired with little finger stimulation; high-pitch sounds with index finger stimulation. For one subject, the stimulus mapping was accidentally reversed. The analysis was adjusted accordingly. **D**, Three-dimensional sound space used for multisensory training (colored dots; 72 sounds per category) and generalization tests/fMRI (black crosses; 72 sounds in total). The relevant dimension (pitch) varied along the x-axis; in the second dimension, which is completely orthogonal to the first, spectral modulation rate was varied in six linear steps; the third dimension (spectral modulation density) consisting of two levels, is only partially indicated to avoid ambiguity. Pitch values used for category training were sampled from two non-overlapping normal distributions with a standard deviation of 1/8 octave. For generalization tests and scanning, we selected six equidistant (log scale) pitch levels between category centers (μ_1 and μ_2). The category boundary was fixed at 200 Hz (f_c).

Experimental procedure. The experiment was distributed over five days, containing two fMRI sessions at the beginning and the end as well as three behavioral training days in between (**Figure 3.2**). Behavioral training and post-training fMRI data acquisition were conducted on four consecutive days; the pre-training fMRI session preceded the behavioral training by maximally ten days. Before data acquisition, participants were familiarized with the ripple sounds and their variability by passively listening to the 72 test sounds while lying in the scanner. To ensure identical stimulation settings for behavioral training and imaging, participants performed the behavioral association training inside the scanner while no images were acquired. They comfortably lied on the scanner bed at accurate viewing distance from the screen. Their head position was not fixed as during actual scanning to provide more comfort and freedom. We employed the same visual, tactile, and auditory stimulation equipment during training and scanning.

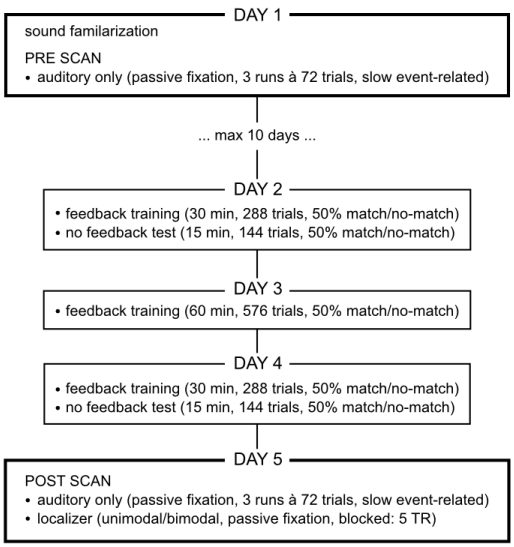


Figure 3.2. Experimental procedure. The behavioral training and scanning procedure was identical for all subjects. Training and post-training fMRI acquisitions were always completed on four consecutive days. All experimental parts were conducted inside the magnet (no image acquisition during behavioral training).

Behavioral training. Category training consisted of a spatial crossmodal association task in which all sounds on either side of the category boundary had to be matched with a visual/tactile stimulus appearing at one of two possible spatial locations (see **Figure 3.1**). For the AV training group, each sound co-occurred with the same grating at one of two possible locations. Analogously, participants from the AT training group felt the same vibratory stimulation on one of two possible fingers on their left hand when hearing a sound. Participants of both groups were instructed to fixate the central fixation cross and to indicate whether the crossmodal pair was a match or no-match by pressing a button with their right index or middle finger within 3000 ms from stimulus onset. To avoid verbalization of the learning rule, no instructions were given to the participants with respect to the relevant sound feature for accurate crossmodal mapping; rather participants implicitly learned the correct association with the help of visual feedback. Their selection was displayed after stimulus offset as a small filled circle (1/2° visual angle) in the screen center, broad-rimmed (1° visual angle) by the correct response after lapse of the response time (see **Figure 3.3A**). Matches were displayed in green; no-matches were displayed in red. Single-color

filled circles thus represented correct responses, while multicolored circles represented a mismatch between the participant's response and the correct response (**Figure 3.3B**). Participants were not informed about the purpose of sound category learning underlying the crossmodal association task.

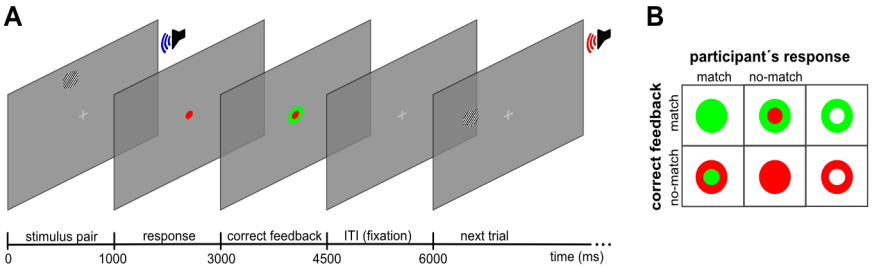


Figure 3.3 Association training procedure. A, Stimulation sequence and timing depicted for an AV association trial. The onset and duration of the sound and the visual grating were matched. Participants had 3000 ms from stimulus onset to respond with a button press whether they considered the stimulus pair a match (green) or no-match (red) until the correct answer was displayed for 1500 ms. **B**, Layout of the feedback procedure. Match/no-match trials were presented with equal probability such that the feedback color as well as button presses are orthogonal to the sound category. The response made by the participant is indicated as a small filled circle in the center of the screen, broad-rimmed by the correct response after lapse of the response time. Single-color filled circles represent a match between the participant's response and the correct response (independent of color). Unfilled circles reflect correct feedback in the case of missing or late response from the participant.

In one feedback training round (288 trials), each of the 144 training sounds was presented in combination with both visual/tactile stimuli in a randomized fashion, creating 50% match and 50% no-match stimulus pairs. Each training round was split into four blocks of 72 trials each, allowing a short break in between. One training round lasted approximately 30 minutes. To assess the generalization of the learned categories to new sounds without the underlying categorical structure, participants performed the same match/no-match task without feedback while presented with the 72 test sounds (black crosses in **Figure 3.1D**), also paired with both visual or tactile stimuli, resulting in 144 trials. This generalization test was performed

after one initial feedback training round (test-day1) and again on the last training day after 3 additional training rounds (test-day3). Hence, test-day1 does not reflect a completely naïve perceptual state but rather an early learning phase after the task procedure is clear to the participants and trial-and-error responses have ceased.

Behavioral data analysis. Reaction times were corrected for outliers (elimination criterion: minimum 1.5 interquartile ranges below the first quartile or at least 1.5 interquartile ranges above the third quartile) and analyzed for correct trials only. We analyzed the within- and between-session learning effects in a mixed-design ANOVA using training group (AV, AT) as a between-subject factor and days (day 1, day 2, day 3) or feedback training blocks (4 blocks for day 1 and day 3, 8 blocks for day 2) as repeated measures. The differences in reaction times and accuracies between the two no-feedback generalization tests were analyzed in a separate model.

Scanning parameters. FMRI data acquisition was performed with a 3-Tesla Siemens Allegra head scanner at the Maastricht Brain Imaging Center (MBIC, Maastricht, the Netherlands). Each subject was scanned twice, once before crossmodal association training and once after completion of four behavioral training rounds (corresponding to a total number of 1152 feedback trials). Both scanning sessions contained three repetitions of the test sound continuum (72 sounds) presented in isolation split into three functional runs of 360 volumes each with additional four volumes of fixation at the beginning of each run. Participants were instructed to fixate and attentively listen to the sounds while functional MRI data were acquired in 37 slices covering almost the whole brain (in most participants only the very anterior tip of the temporal lobe was not covered by the slices). The passive fixation task avoids undesired interference by task-related processes (Foxe et al., 2002; Schürmann et al., 2006; van der Linden et al., 2011) and has been proven to be particularly useful for the investigation of low-level reorganization of sound representations (Ley et al., 2012). Blood-oxygen-level-dependent (BOLD) contrast was measured with a standard echo-planar

imaging sequence in a slow event-related design (repetition time (TR) = 3500 ms; acquisition time (TA) = 2400 ms; field of view (FoV) = 224 x 224; matrix size = 96 x 96; echo time (TE) = 30 ms; voxel dimensions = 2.5 mm isotropic, 10% slice gap). The average intertrial interval between two stimuli was 17.5 s (jittered between 4, 5, and 6 TR). Sounds were presented in silence between volume acquisitions leaving a 50 ms gap before and after. Stimulus order was randomized for each run and subject; stimulus delivery was synchronized with MR pulses using Presentation software (Neurobehavioralsystems). For subsequent overlay and cortex reconstruction purposes, high-resolution T1-weighted structural images (voxel dimensions, 1 mm isotropic) were acquired with an ADNI MPRAGE sequence (192 sagittal slices (whole brain), TR = 2050 ms; TE = 2.6 ms).

The post-training scan session additionally contained a functional localizer for the ripple sounds, visual gratings, and tactile stimuli. The stimulus design contained seven conditions, three unimodal (visual-only, auditory-only, and tactile-only) and four bimodal (audio-visual congruent and audio-visual incongruent, audio-tactile congruent and audio-tactile incongruent), each presented in six blocks of five stimuli each. In half of the unimodal visual and tactile blocks, the stimulus was presented in spatial location 1 (i.e. 110º or little finger) and in the other half in spatial location 2 (i.e. 160º or index finger). Only the category centers (μ_A and μ_B) and two neighboring pitches (μ_{A-1}, μ_{A+1}; μ_{B-1}, μ_{B+1}) were used for the six auditory blocks. To cover the large feature variability, different densities and velocities were used for the five stimuli within one block. The scanner parameters (TR, TA, TE, FoV, matrix and voxel dimensions as well as slice positioning) corresponded to the main experiment; experimental blocks contained five TR (one stimulus per TR in acquisition gap) interspersed with baseline blocks of four TR length (14 s). In total 378 volumes (experimental: 7 conditions x 6 blocks x 5 volumes = 210; baseline: 7 conditions x 6 blocks x 4 volumes = 168) were acquired plus four additional TR of fixation.

FMRI data analysis. Initial data analyses were performed with BrainVoyager QX (Brain Innovations). The first four volumes of each run (including the

localizer) were discarded from the analysis to allow for T1 equilibrium. Preprocessing of functional images included slice scan time correction (cubic spline interpolation), 3D motion correction (trilinear/sinc interpolation) and temporal high-pass filtering (GLM with Fourier basis set (includes linear trend), 2 cycles). The six individual motion parameters from the rigid body transformations (three translation and three rotation parameters) were included as confound predictors in the design matrix after normalization. Moderate spatial smoothing (FWHM: 4 mm) was applied to the localizer data only. Functional scans were aligned to the anatomical datasets and both functional and anatomical data were normalized to Talairach space (Talairach and Tournoux, 1988). To allow the visualization of activation maps on the surface, grey- and white matter were segmented and the curvature information was used to reconstruct a 3D mesh representation. To increase the functional-anatomical correspondence across participants, the individual surface reconstructions were aligned using a moving target group average (cortex-based alignment, CBA, Goebel et al., 2006). The information from this conversion was used during functional group analyses.

Except for the analysis of congruency effects, data from both training groups were merged for the analysis of the functional localizer. We computed a general linear model (GLM) by fitting the BOLD response time course with the predicted time series for the seven experimental conditions. The hemodynamic response delay was corrected for by convolving the predicted time courses with a canonical (double gamma) hemodynamic response function. For three of the participants from the AV training group the localizer contained no tactile conditions due to malfunction of the piezostimulator. Functional data from one subject (AV group) was entirely discarded due to excessive head motion (> 3.5 mm) preventing adequate alignment of pre- and post-training fMRI sessions. We performed group (fixed-effects) analyses of the main unimodal contrasts (visual > baseline), (tactile > baseline), and (auditory > baseline) as well as stringent contrasts to reveal regions specifically recruited during the processing of the bimodal stimuli defined by ((auditory > baseline) ∩ (visual > baseline) ∩ (audio-

visual > auditory) ∩ (audio-visual > visual)) as well as ((auditory > baseline) ∩ (tactile > baseline) ∩ (audio-tactile > auditory) ∩ (audio-tactile > tactile)). Furthermore, we were interested in regions of multisensory convergence, i.e. processing both sounds as well as the visual/tactile stimuli in isolation. For this, we calculated the conjunctions of the unimodal contrasts ((auditory > baseline) ∩ (tactile > baseline)) and ((auditory > baseline) ∩ (visual > baseline)). To reveal specific learning-induced multisensory congruency effects, we defined the conjunction between the main effects of audio-visual congruency (incongruent > congruent) for the AV group and the interaction with training group. The piezoelectric system failure for three participants from the AV group created unequal sample sizes for the analog audio-tactile congruency contrast, which is why we tested the main effect of audio-tactile congruency for the AT group without the group interaction.

In the analysis of the main experiment, predictors were pooled over all test sounds to investigate potential differences in responses to sounds after AV versus AT association training. We defined a fixed-effects group contrast to identify regions that show an increase in activation from the pre-training to the post-training scan ((post > pre) ∩ (post > baseline)) specifically for one training group and not the other (AV > AT) or (AT > AV). The conjunction of these contrasts controls for potential differences in the pre-session and training-independent differences between sessions.

Thresholds for contrast maps were corrected for multiple comparisons on a whole brain basis using a cluster-size estimation procedure (Forman et al., 1995), which takes the spatial neighborhood relationships into account when calculating the error probability (Goebel et al., 2006). We performed 1000 iterations of Monte Carlo simulations estimating the frequencies of different cluster sizes in random noise-only activation maps; the initial uncorrected threshold was set to $p < .05$ unless indicated otherwise. All maps are displayed with a minimum cluster size threshold, which yields a false-positive rate of 5% and are noted together with the respective minimum cluster size (cs).

3 | RESULTS

Behavioral results. Inspection of average performance plotted across days (**Figure 3.4A**) reveals large similarities between the behavioral learning effects in the two training groups. In both groups, the learning curves depict quick improvements at early learning stages (mainly day 1) reaching an asymptotic level on the last training day. This is reflected in the significant main effect of training days for reaction times ($F_{2,24}$ = 35.2, p < .001) and accuracies ($F_{2,24}$ = 29.7, p < .001), which were not qualified by an interaction with training group. Post-hoc pairwise comparisons for accuracy revealed significant differences between day1 and day2 (p < .001) as well as day2 and day3 (p = .015). Reaction times settled a bit faster, showing large differences between day1 and day2 (p < .001) and only marginally significant differences between day2 and day3 (p = .068).

A significant difference in reaction times between training groups could only be attested for the first training day (t_{12} = 2.2, p = .049). Participants from the AT training group had a slower initial performance: mean reaction time on the first day: 1551 ms (AT) and 1279 ms (AV). The AT group reaches an asymptotic accuracy of 93% averaged across blocks on the last training day while the AV group marginally lags behind with an average performance of 88% (difference not significant, p = .34).

Apart from the between-session learning effect, in both groups reaction times significantly dropped across blocks within day 1 (blocks 1 - 4) ($F_{3,36}$ = 7.3, p = .001, shaded areas in **Figure 3.4A**), while accuracies significantly increased across blocks on day 1 ($F_{3,36}$ = 9.4, p < .001) and also still on day 2 (blocks 6 - 13) ($F_{7,84}$ = 2.4, p = .027). **Figure 3.4B** shows that participants display a large amount of generalization of the association rule, reflected in the significant decrease in reaction times from the first to the second test session ($F_{1,12}$ = 8.2, p = .014), and the significant increase in accuracies ($F_{1,12}$ = 10.5, p = .007) despite the lack of feedback or underlying categorical structure (i.e. the pitch values of the test sounds were sampled from a continuum between category means) as well as the novelty of the test sounds, which were not used for training. Due to rather large variability

within groups, there is no significant difference between groups for accuracies or reaction times on test-day1 and test-day2.

FMRI results - functional localizer. The three unimodal contrasts revealed distinct activation maps for processing visual gratings, ripple sounds, as well as vibrotactile stimuli (**Figure 3.5**). The simple visual gratings, presented at two proximal positions (distance: 50º) in the upper-left quadrant of the

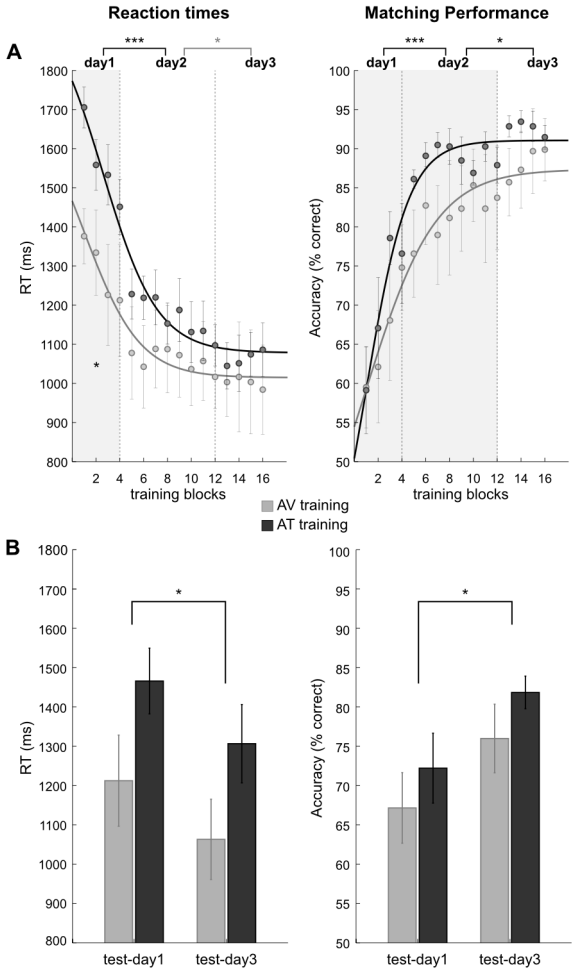


Figure 3.4 Behavioral data during feedback training (A) and generalization tests (B) for the AV and AT training groups. A, Mean (± *SEM*) reaction times (left) and accuracies (right) in blocks of 72 trials across three consecutive training days fitted with a logistic function. The shaded area marks the time windows of significant ($p < .05$) within-session performance change, independent of group. Both, reaction times and accuracies are characterized by significant (*** $p < .001$, * $p < .05$) changes across days. The asterisk in the reaction time plot on the left reflects significant group differences on day 1. **B**, Generalization tests (144 trials) were performed after four training blocks (test-day1) and at the end of the last training day after additional 12 training blocks (test-day3) without corrective feedback using the sounds from the pitch continuum between category centers. Data represent mean (± *SEM*) reaction times (left) and accuracies (right).

visual field elicited widespread activation of a bilateral fronto-parietal visual processing network (**Figure 3.5A**). The gratings evoked significant activation in superior parietal cortex (anterior and posterior portions of the intraparietal sulcus (IPS)) on both hemispheres with a tendency for stronger activation in the right compared to the left hemisphere. Furthermore, activation patterns involved the junction of the superior frontal sulcus (SFS) and the precentral sulcus, most likely corresponding to the human frontal eye fields (FEF) as well as a posterior portion of the inferior frontal sulcus (IFS, intersection with the precentral sulcus). In addition, BOLD signal changes in response to the visual gratings can be observed in lateral occipito-temporal areas, most likely corresponding to the human homologue of the MT complex (V5, hMT+), as well as ventral portions of the occipito-temporal complex (VOT). In the right hemisphere activation patterns extended into the inferior portion of the occipito-temporal regions including the occipito-temporal sulcus and parts of the fusiform gyrus. The relatively short duration of the visual stimulus presentation as well as the limited number of trials (15 per spatial location) did not allow the demarcation of retinotopic locations in the primary visual cortex.

As can be seen in **Figure 3.5B**, ripples activated a large expanse of the superior temporal cortex on the superior temporal gyrus (STG), including Heschl's gyrus (HG) and the posteriorly adjacent Heschl's sulcus (HS) as well as planum temporale (PT). Anteriorly, activation spreads into the first transverse sulcus (FTS) and the planum polare (PP).

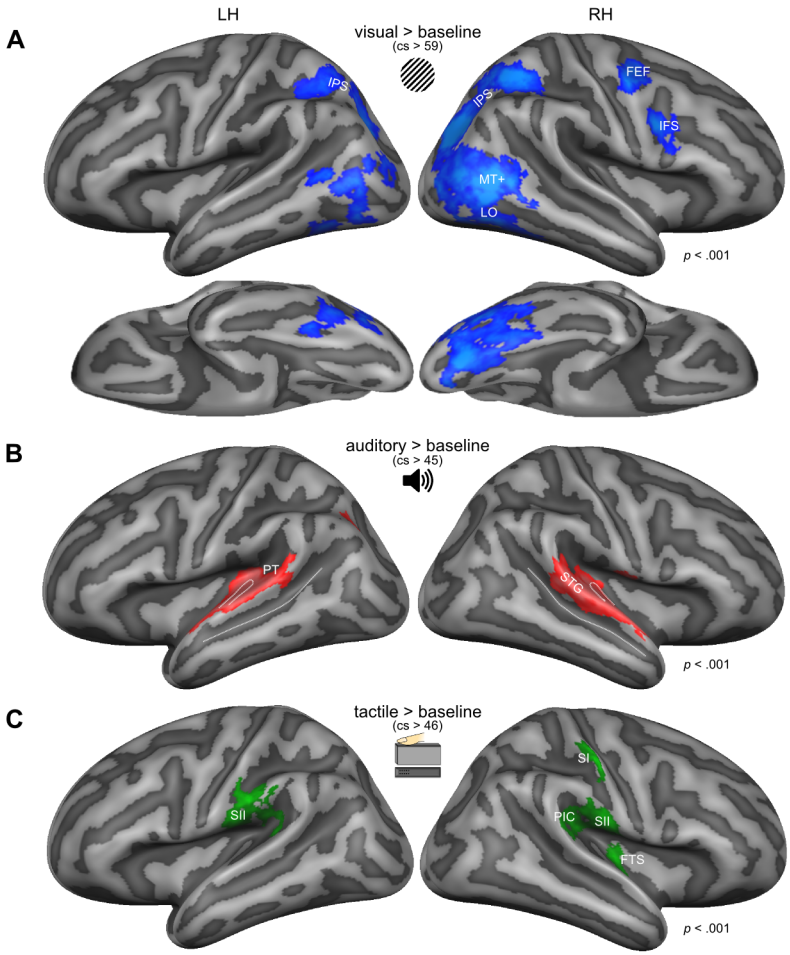


Figure 3.5 Activation maps from the localizer scan. Data from both groups (N = 13) were combined to obtain training-independent activations of **A** visual gratings, **B** ripple sounds, and **C** vibrotactile finger stimulation. Data are shown on a group-average cortex reconstruction. Heschl's gyrus (HG) and superior temporal sulcus (STS) are demarcated in B for better orientation. LH, left hemisphere; RH, right hemisphere; IPS, intraparietal sulcus; FEF, frontal eye fields; IFS, inferior frontal sulcus; LO, lateral occipital complex; MT+, human motion complex; PT, planum temporale; STG, superior temporal gyrus; SI, primary somatosensory cortex; SII, secondary somatosensory cortex; FTS, first transverse sulcus; PIC, posterior insular cortex.

As expected, vibratory stimulation of the index and little finger of the left hand activated the primary somatosensory area (SI) corresponding to the hand area in the middle of the right postcentral gyrus (**Figure 3.5C**). In addition, prominent activity clusters could be observed in a secondary somatosensory region (SII) in the inferior parietal lobe on the ceiling of the lateral sulcus bilaterally. Moreover, tactile stimulation evoked significant activation in the posterior temporal lobe at the posterior end of the lateral sulcus, most likely corresponding to the posterior insular cortex (PIC) as well as a strip in the medial parts of the lateral sulcus, in the FTS on the right but not the left hemisphere.

The intersection of the unimodal auditory and unimodal visual contrasts revealed one significant cluster in the left anterior IPS close to the postcentral sulcus, probably corresponding to IPS5 (Silver & Kastner, 2009). The analogous contrast for auditory and tactile stimuli showed a large expanse of regions in the superior temporal lobe (**Figure 3.6A**). Areas of audio-tactile co-activation lie on the superior rim of posterior STG, extending into the PT bilaterally as well as medial parts, spreading along the FTS in the right hemisphere, extending into the lateral sulcus. Normalized model coefficients (betas) suggest audio-tactile specificity of these temporal activation sites while the parietal region is driven by all three modalities but to a lesser degree (**Figure 3.6B**). The BOLD signal time course corresponding to the posterior part of the auditory cortex (PT) reflects a benefit from the combined presentation of sounds and vibrotactile stimuli in contrast to unimodal stimulation (**Figure 3.6C**). These areas show significantly increased BOLD signal change for multimodal compared to either unimodal stimuli, during the entire block duration. No significant activation clusters could be identified for the analogous contrast for auditory and visual stimuli.

No group-specific learning-induced congruency effects for the audio-visual stimuli could be revealed. While unequal group sizes did not allow to investigate the interaction of audio-tactile congruency and training group, no significant congruency effects could be found for the AT group alone.

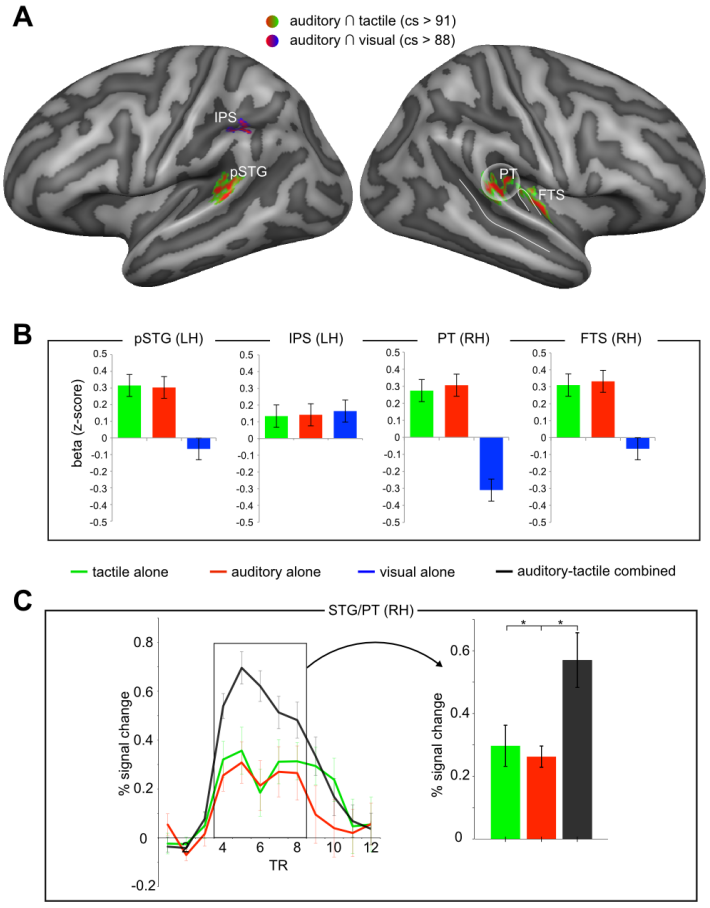


Figure 3.6 Sites of multisensory convergence and interaction. A, Group (N = 13) activation maps of the conjunctions of auditory and visual activations as well as auditory and tactile activations. **B**, Corresponding model coefficients. **C**, Average time course of the posterior STG on the right hemisphere (see encircled areas in A) during entire block duration. The black rectangle indicates the time window averaged on the right. For corresponding Talairach coordinates, see Table 3.1.

FMRI results - main experiment. Despite identical stimulation conditions in the post-learning fMRI session, sound activation patterns for the AV group significantly differed from the AT group after training. **Figure 3.7A** illustrates those regions exhibiting significant ($p < .05$, cluster-size corrected) group-specific training effects. AT training significantly increased the responses to ripple sounds in medial areas of the right temporal lobe including the PIC as

well as a region in the posterior belt of the primary auditory cortex, most likely corresponding to area CM (caudomedial auditory belt area), described in studies of animal electrophysiology (Schroeder et al., 2001; Fu et al., 2003). Smaller clusters were found on the superior rim of middle STG on the left. In contrast, listening to the same sounds after three days of AV training resulted in a significant increase in activation in a fronto-parietal processing network, predominantly on the left hemisphere including anterior and posterior portions of the IPS as well as FEF. Furthermore, the anterior insular cortex (AIC) on both hemispheres as well as the left anterior cingulate cortex (ACC) were more strongly engaged after AV training than before. Interestingly, these training-sensitive regions partially overlapped with purely visual or tactile processing areas, identified during the localizer scan. As reflected by the time course (**Figure 3.7B**), the anterior and posterior regions in the IPS did not respond to sounds prior to association training but are significantly recruited by the same sounds after training in the AV group only. Correspondingly, ripple responses in the right PIC significantly increase with AT but not with AV association learning.

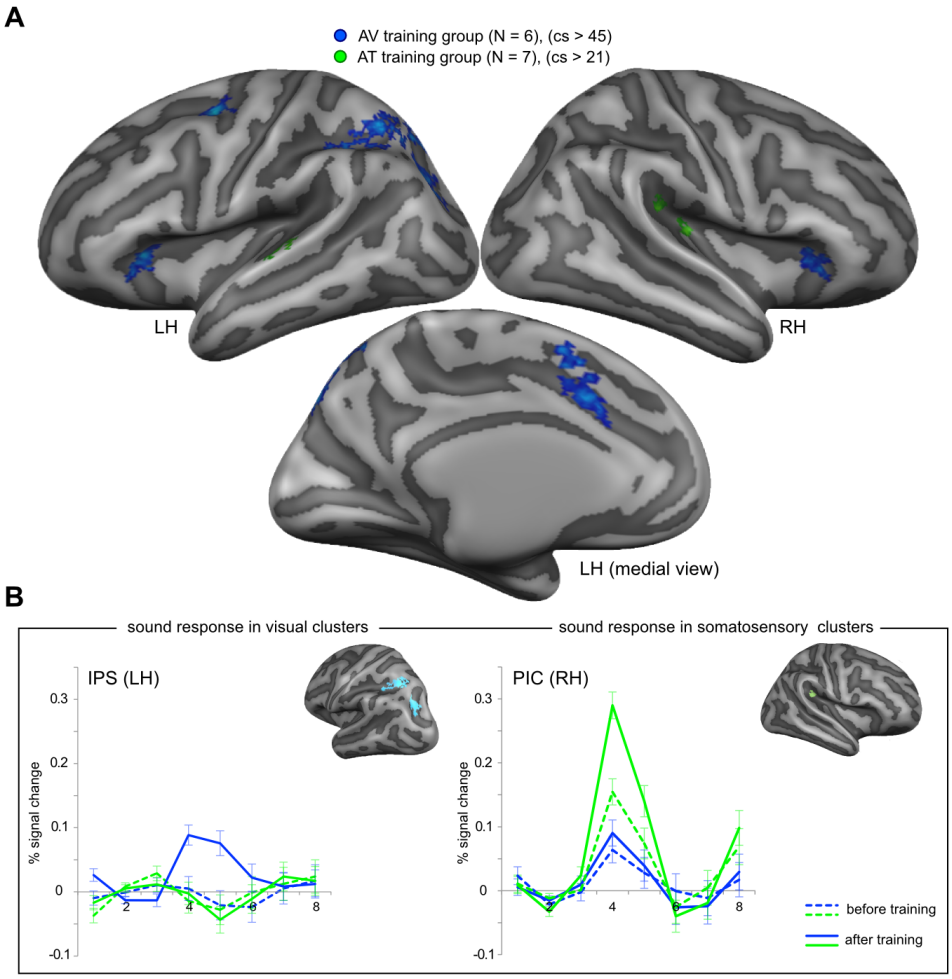


Figure 3.7 Group-specific training effects. A, Cortical regions showing a learning-specific increase in activation to the same ripple sounds. **B**, Average time course in those regions overlapping with the visual (left) and tactile (right) clusters identified in the separate localizer scan for the AV (blue) and AT (green) training group. For Talairach coordinates, see Table 3.1.

Contrast	ROI	hemisphere	x	y	z
A ∩ V	IPS	LH	-43	-43	41
A ∩ T	pSTG	LH/RH	-52/58	-38/-32	17
	PT	RH	39	-31	19
	FTS	RH	37	20	7
AV > AT ∩ post > pre	AIC	LH/RH	-28/29	23/23	9
	FEF	LH	-24	-6	46
	ACC	LH	-11	11	37
	SPL	LH	-12	-71	47
	pIPS	LH	-27	-66	24
	aIPS	LH	-28	-53	42
AT > AT ∩ post > pre	STG	LH	-57	-25	11
	CM	RH	33	-24	16
	PIC	RH	45	-32	24

Table 3.1 Talairach coordinates (x, y, z) of group average activation maps for bimodal activation sites (Figure 3.6) and group-specific training contrasts (post > pre, see Figure 3.7). ROI, region of interest. Coordinates correspond to ROI centroid.

4 | DISCUSSION

In this study, we combined two different multisensory training paradigms with fMRI to investigate experience-dependent crossmodal plasticity. Using a controlled design with identical acoustic ripple stimuli, we revealed

striking differences in learning-induced changes in sound responses with audio-visual versus audio-tactile association training. After audio-visual training, sounds recruited regions in posterior parietal cortex (PPC) that before training did not respond to auditory stimuli. Conversely, audio-tactile learning increased sound processing in auditory cortex and posteriorly adjacent regions. Our data suggest that multisensory training promotes the transformation from an originally acoustic sound representation to an experience-dependent multimodal representation, which persists even without bottom-up visual or tactile information. The discrepancy between the learning-induced changes for the two training groups corroborates earlier findings on profound differences in the underlying mechanisms of audio-visual and audio-tactile interaction (Schroeder et al., 2001; Schroeder and Foxe, 2002) that might promote crossmodal reorganization at different neural processing levels.

Learning-induced reorganization of sound responses

Audio-visual association training significantly increased sound responses in large parts of a dorsal fronto-parietal network, previously described in the context of auditory representations of spatial location (Renier et al., 2009), top-down directed mechanisms of attention (Corbetta & Shulman, 2002), as well as visual spatial imagery (Formisano et al., 2002; Trojano et al., 2000; Sack, 2009). After audio-visual training (relative to pre-training activation levels and AT training effects), passive listening to the isolated sounds strongly engaged the superior parietal cortex, which has been shown to play a role in audio-visual integration (Molholm et al., 2006) as well as auditory, visual, and crossmodal spatial attention (Bushara et al., 1999; Szczepanski, Konen, & Kastner, 2010). Furthermore, we observed a significant effect of audio-visual learning on bilateral AIC together with the FEF and ACC, forming a processing network commonly activated during perceptually demanding tasks involving crossmodal information exchange and binding (Sterzer & Kleinschmidt, 2010). In this respect, it is conceivable that the ripple sounds elicited processes of crossmodal spatial attention targeting

representations of the associated visual component for binding of the multimodal object features previously encountered during training. This interpretation is also in line with memory and imagery studies, which demonstrated fundamental overlap of cortical regions involved in perception as well as recall or imagery of the same items (Nyberg et al., 2000; Trojano et al., 2000; Wheeler et al., 2000). The functional diversity of the fronto-parietal network, reported in earlier studies, complicates the definition of the exact mechanisms underlying its learning-induced sound response in the current experimental context unequivocally. Yet, it is striking that simple ripple sounds activated regions of complex higher-order processing functions instead of early visual or multisensory cortex according to earlier reports of audio-visual crossmodal effects. This reflects the multifaceted nature of crossmodal plasticity, which seems to depend on the employed stimulus material and the experimental task and learning design. The crossmodal recruitment of primary and early visual areas (Mcintosh et al., 1998; Zangenehpour & Zatorre, 2010) seems to rely on intrinsic correspondences between the auditory and visual stimuli and a high level of crossmodal expectancy. This is not provided in the current experimental context given the complex many-to-one association task and the lack of visual stimuli during the post-learning fMRI session. Furthermore, crossmodal activations in typical multimodal areas in the superior temporal sulcus (STS) are common for natural object categories (Beauchamp et al., 2004; van der Linden et al., 2011) and have only been observed for arbitrary audio-visual stimulus pairs during active association tasks or conditioning paradigms (Meyer et al., 2007), decreasing with the level of expertise (Tanabe et al., 2005). Thus, it is possible that only active crossmodal processes would recruit semantic object representations in these regions.

In contrast to the involvement of higher-order fronto-parietal processing regions following audio-visual learning, the representational changes in the AT group were restricted to the temporal lobe, particularly medial and posterior portions of the Sylvian fissure. This indicates that the targets of

learning-induced reorganization lie closer to classical auditory processing regions in the case of audio-tactile interaction, supporting views of early audio-tactile convergence and interaction in the auditory cortex (Foxe & Schroeder, 2005; Ghazanfar & Schroeder, 2006; Musacchia & Schroeder, 2009). Interestingly, audio-tactile association learning caused enhanced processing of ripple sounds in a medial portion of the posterior auditory belt, possibly corresponding to area CM, identified in monkeys as site of receiving somatosensory input (Schroeder et al., 2001; Fu et al., 2003). This suggests that post-learning sound processing involves regions coding somatosensory information relevant for sound identification even in absence of bottom-up somatosensory input. Corresponding to the observations from the audio-visual learning group, sounds also activated areas of non-auditory preference, such as the PIC after AT training. Although not directly part of the somatosensory cortex, the PIC exhibits a variety of somatosensory processing properties (Craig et al., 2000) and represents an important hub for crossmodal information exchange due to its particularly advantageous anatomical position between auditory and somatosensory cortices (Augustine, 1996). It is highly interconnected with the secondary somatosensory and auditory cortex and might be the source of somatosensory activation in the auditory cortex (Hackett et al., 2007).

The discrepancy between the involvement of low-level auditory areas and a high-level fronto-parietal network for the AT and AV group respectively points towards crucial differences in the underlying mechanisms of crossmodal interaction. Evidence from electrophysiology suggests that somatosensory input converges with auditory input at early levels of the auditory processing hierarchy in area CM of the auditory association cortex in a feed-forward fashion (Schroeder et al., 2001), while visual responses in the same area are characterized by feedback profiles (Schroeder & Foxe, 2002). Hence, it could be speculated that audio-visual plasticity might be mediated via feedback projections to the auditory cortex from higher-order association cortices in the superior parietal and frontal lobe, while audio-tactile interaction might be based on sensory processes via direct

connectivity of somatosensory and auditory cortices (Driver & Noesselt, 2008; Klemen & Chambers, 2012). Obviously, functional connectivity analyses (preferably during learning) are recommended to verify this interpretation. Interestingly, reaction times and accuracies during behavioral training were insensitive to potential differences in crossmodal interaction. Both groups learned the tasks equally well, reflected in their asymptotic learning curves and good performance level on the last day.

Crossmodal convergence and plasticity

Independent of training, we identified a cluster of multimodal convergence in the anterior IPS, which responded equally well to visual gratings, vibrotactile finger stimulation and acoustic ripples. These polymodal characteristics of the parietal cortex are consistent with previous findings (Bremmer et al., 2001; Grefkes et al., 2002), providing evidence for the modulatory role of the parietal association cortex in crossmodal information exchange (Culham & Kanwisher, 2001; Klemen & Chambers, 2012). Crucially, this multimodal region remained unaffected by learning. Instead, after AV learning, sounds recruited more posterior regions in the IPS specifically coding the visual component of the audio-visual objects, previously unresponsive to auditory stimuli alone, indicative of experience dependent crossmodal plasticity. This reflects similarities to crossmodal recruitments in blindfolded participants (Pascual-Leone et al., 2005) and further supports the mounting evidence for the multisensory nature and general plasticity of the cortex (Ghazanfar & Schroeder, 2006) where processing resources are adapted to the environmental demands, especially in cases of unreliable and ambiguous information from one modality. The spatial dissociation of these processes in the IPS is in line with evidence for diverse functional properties of separate subregions within the PPC (Culham & Kanwisher, 2001; Szczepanski et al., 2010). It also seems that learning-induced activation increases spared regions of audio-tactile convergence in the superior rim of the STG. A variety of regions in the auditory cortex, especially posteriorly as well as anterior-medially of HG, show a clear

preference for auditory as well as somatosensory input while visual input is slightly suppressed. The clusters in the right posterior STG and PT additionally show signs for multisensory interaction: their response significantly increases when both, auditory and somatosensory stimuli are presented together, exceeding the response to each individual stimulus alone. Following the ongoing debate about valid criteria for BOLD-related responses qualifying as multisensory integration (Calvert, 2001; Beauchamp, 2005; Goebel & van Atteveldt, 2009), we refrain from labeling these areas accordingly, as we cannot tell apart the possible underlying mechanisms leading to the observed response profile. The data from our localizer are in line with mounting evidence from human fMRI studies for tactile influences and modulations in the auditory belt region (Foxe et al., 2002; Schürmann et al., 2006).

In conclusion, this study revealed that crossmodal spatial association training has lasting effects on subsequent unimodal sound processing. The comparison of audio-visual with audio-tactile learning paradigms allowed investigating the specifics of crossmodal plasticity with a focus on different levels of multisensory interaction. The precise functional role of the learning-induced changes in sound representations remains to be investigated in future studies. In the context of the current experiment, it would be of particular interest to test whether the newly formed multisensory representations contain specific information enabling the decoding of the newly learned pitch categories. This requires advanced analyses techniques with higher sensitivity to spatially distributed activation patterns.

REFERENCES

Augustine JR (1996) Circuitry and functional aspects of the insular lobe in primates including humans. Brain Res Rev 22: 229–244.

Beauchamp MS (2005) Statistical criteria in fMRI studies of multisensory integration. Neuroinformatics 3: 93–113.

Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41: 809–823.

Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann K, Zilles K, Fink GR (2001) Polymodal motion processing in posterior parietal and premotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys. Neuron 29: 287–296.

Bushara KO, Weeks RA, Ishii K, Catalan M-J, Tian B, Rauschecker JP, Hallett M (1999) Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. Nat Neurosci 2: 759–766.

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11: 1110–1123.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. Science 276: 593–596.

Chi T, Gao Y, Guyton MC, Ru P, Shamma S (1999) Spectro-temporal modulation transfer functions and speech intelligibility. J Acoust Soc Am 106: 2719–2732.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3: 201–215.

Craig AD, Chen K, Bandy D, Reiman EM (2000) Thermosensory activation of insular cortex. Nat Neurosci 3: 184–190.

Culham JC, Kanwisher NG (2001) Neuroimaging of cognitive functions in human parietal cortex. Curr Opin Neurobiol 11: 157–163.

Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. Neuron 57: 11–23.

Driver J, Spence C (1998) Attention and the crossmodal construction of space. Trends Cogn Sci 2: 254–262.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic

resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 33: 636–647.

Formisano E, Linden DEJ, Salle F Di, Trojano L, Esposito F, Sack AT, Grossi D, Zanella FE, Goebel R (2002) Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. Neuron 35: 185–194.

Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. Neuroreport 16: 419–423.

Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM (2002) Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J Neurophysiol 88: 540–543.

Fu K-MG, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE (2003) Auditory cortical neurons respond to somatosensory stimulation. J Neurosci 23: 7510–7515.

Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10: 278–285.

Goebel R, Van Atteveldt N (2009) Multisensory functional magnetic resonance imaging: a future perspective. Exp Brain Res 198: 153–164.

Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum Brain Mapp 27: 392–401.

Grefkes C, Weiss PH, Zilles K, Fink GR (2002) Crossmodal processing of object features in human anterior intraparietal cortex: An fMRI study implies equivalencies between humans and monkeys. Neuron 35: 173–184.

Hackett TA, Smiley JF, Ulbert I, Karmos G, Lakatos P, De la Mothe LA, Schroeder CE (2007) Sources of somatosensory input to the caudal belt areas of auditory cortex. Perception 36: 1419–1430.

Klemen J, Chambers CD (2012) Current perspectives and methods in studying neural mechanisms of multisensory interactions. Neurosci Biobehav R 36: 111–133.

Kowalski N, Depireux DA, Shamma SA (1996) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J Neurophysiol 76: 3503–3523.

Von Kriegstein K, Giraud A-L (2006) Implicit multisensory associations influence voice recognition. PLoS Biol 4: e326.

Von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud A-L (2005) Interaction of face and voice areas during speaker recognition. J Cogn Neurosci 17: 367–376.

Ley A, Vroomen J, Hausfeld L, Valente G, De Weerd P, Formisano E (2012) Learning of new sound categories shapes neural response patterns in human auditory cortex. J Neurosci 32: 13273–13280.

Van der Linden M, Van Turennout M, Fernández G (2011) Category training induces cross-modal object representations in the adult human brain. J Cogn Neurosci 23: 1315–1331.

Mcintosh AR, Cabeza RE, Lobaugh NJ (1998) Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. J Neurophysiol 80: 2790–2796.

Meyer M, Baumann S, Marchina S, Jancke L (2007) Hemodynamic responses in human multisensory and auditory association cortex to purely visual stimulation. BMC Neurosci 8: 14.

Molholm S, Sehatpour P, Mehta AD, Shpaner M, Gomez-Ramirez M, Ortigue S, Dyke JP, Schwartz TH, Foxe JJ (2006) Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. J Neurophysiol 96: 721–729.

Murray MM, Foxe JJ, Wylie GR (2005) The brain uses single-trial multisensory memories to discriminate without awareness. NeuroImage 27: 473–478.

Musacchia G, Schroeder CE (2009) Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. Hear Res 258: 72–79.

Nyberg L, Habib R, McIntosh AR, Tulving E (2000) Reactivation of encoding-related brain activity during memory retrieval. PNAS 97: 11120–11124.

Ohl FW, Scheich H, Freeman WJ (2001) Change in pattern of ongoing cortical activity with auditory category learning. Nature 412: 733–736.

Pascual-Leone A, Amedi A, Fregni F, Merabet LB (2005) The plastic human brain cortex. Annu Rev Neurosci 28: 377–401.

Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, Sams M (2005) Primary auditory cortex activation by visual speech: an fMRI study at 3T. Neuroreport 16: 125–128.

Renier LA, Anurova I, De Volder AG, Carlson S, VanMeter J, Rauschecker JP (2009) Multisensory integration of sounds and vibrotactile stimuli in processing streams for “what” and “where”. J Neurosci 29: 10950–10960.

Sack AT (2009) Parietal cortex and spatial cognition. Behav Brain Res 202: 153–161.

Scheich H, Brechmann A, Brosch M, Budinger E, Ohl FW, Selezneva E, Stark H, Tischmeyer W, Wetzel W (2011) Behavioral semantics of learning and crossmodal processing in auditory cortex: the semantic processor concept. Hear Res 271: 3–15.

Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cogn Brain Res 14: 187–198.

Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC (2001) Somatosensory input to auditory association cortex in the macaque monkey. J Neurophysiol 85: 1322–1327.

Schürmann M, Caetano G, Hlushchuk Y, Jousmäki V, Hari R (2006) Touch activates human auditory cortex. NeuroImage 30: 1325–1331.

Schürmann M, Caetano G, Jousmäki V, Hari R (2004) Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. J Acoust Soc Am 115: 830.

Seitz AR, Kim R, Shams L (2006) Sound facilitates visual learning. Curr Biol 16: 1422–1427.

Shams L, Seitz AR (2008) Benefits of multisensory learning. Trends Cogn Sci 12: 411–417.

Silver MA, Kastner S (2009) Topographic maps in human frontal and parietal cortex. Trends Cogn Sci 13: 488–495.

Sterzer P, Kleinschmidt A (2010) Anterior insula activations in perceptual paradigms: often observed but barely understood. Brain Struct Funct 214: 611–622.

Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. J Acoust Soc Am 26: 212–215.

Szczepanski SM, Konen CS, Kastner S (2010) Mechanisms of spatial attention control in frontal and parietal cortex. J Neurosci 30: 148–160.

Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme.

Tanabe HC, Honda M, Sadato N (2005) Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. J Neurosci 25: 6409–6418.

Trojano L, Grossi D, Linden DEJ, Formisano E, Hacker H, Zanella FE, Goebel R, Di Salle F (2000) Matching two imagined clocks: the functional anatomy of

spatial analysis in the absence of visual stimulation. Cereb Cortex 10: 473–481.

Wheeler ME, Petersen SE, Buckner RL (2000) Memory's echo: Vivid remembering reactivates sensory-specific cortex. PNAS 97: 11125–11129.

Zangenehpour S, Zatorre RJ (2010) Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. Neuropsychologia 48: 591–600.

Chapter 4

LEARNING OF NEW SOUND CATEGORIES: DISTINCT NEURAL EFFECTS OF VISUAL AND TACTILE CONTEXTUAL INFORMATION

Ley, A., Hausfeld, L., Kaas, A., Vroomen, J., Formisano, E. (in preparation). Learning of new sound categories: distinct neural effects of vision and tactile contextual information.

That is what learning is.
You suddenly understand something
you've understood all your life, but in a new way

~ Doris Lessing

ABSTRACT

Mounting evidence suggests that neural sound representations are no static mappings of the acoustic sound features but prone to dynamic changes according to the environmental demands and the goal of the listener (Ohl et al., 2001; Ley et al., 2012). Despite strong indications for low-level input from the visual and somatosensory modality in the auditory belt (Foxe et al., 2002; Schürmann et al., 2006; Kayser et al., 2009), the effect of crossmodal modulation on rapid learning-induced plasticity remains largely unexplored. Therefore, in the current study, two groups of subjects underwent matched audio-visual or audio-tactile spatial association training, which indirectly induced the formation of distinct pitch categories. Responses to a novel pitch continuum were measured prior to and following three days of association training during passive listening to the sounds in isolation by means of functional magnetic resonance imaging (fMRI). Using multi-voxel pattern analyses (MVPA), we aimed to discern the neural representations of learning-induced perceptual categories. Our results show that only after audio-tactile training, abstract pitch categories could successfully be decoded from locally distributed activation patterns in the auditory cortex and the posterior lateral sulcus, overlapping with somatosensory regions. These findings corroborate the view that the auditory cortex serves complex perceptual operations and indicate that the close feature correspondence of auditory and vibrotactile stimuli induced a functional coupling of the respective stimulus representations during learning, persisting even in absence of tactile stimuli during passive listening. According to electrophysiological recordings, tactile input reaches the auditory belt through forward connections, whereas visual signals are relayed via higher-order multisensory integration areas (Schroeder & Foxe, 2002; Hackett et al., 2007). This discrepancy could potentially account for the superior effect of audio-tactile training on category discrimination in the auditory cortex observed in the current experiment.

1 | INTRODUCTION

To optimally exploit the limited processing capacity of the auditory system, neural sound representations need to remain malleable to the demands of the complex and ever-changing natural acoustic environment. Evidence for this premise has been provided by auditory learning studies in animals (Ohl et al., 2001) and humans (Ley et al., 2012). These studies showed that learning to group artificial sounds into behaviorally meaningful categories causes rapid reorganization of the neural representations in the auditory cortex away from the explicit mapping of the acoustic sound features to a functionally relevant abstract sound representation. Natural sound categories are typically characterized by large acoustic variability within as well as between categories (Harnad, 1987). Extracting the acoustic features relevant for categorization therefore requires the integration of contextual information. Although natural environments are essentially multisensory (Stein and Meredith, 1993), little attention has been paid to the influence of contextual cues from the visual or tactile domain on learning-induced plasticity of sound representations (Russ et al., 2007). Investigating the effect of multisensory interaction during category learning on the formation of categorical sound representations is of particular interest with respect to the ongoing debate about the level of the auditory processing hierarchy at which categorical processing is accomplished. While one line of evidence points towards a generic role of the left superior temporal sulcus (STS) for categorical sound representations (Leech et al., 2009; Liebenthal et al., 2010), recent studies mark the involvement of low-level sensory cortex (Ley et al., 2012) as well as higher-order processing regions in the prefrontal cortex (PFC) in categorical processing (Husain et al., 2006; Lee et al., 2012; Myers & Swan, 2012). As these studies largely differ with respect to the employed stimulus material (ranging from artificial ripples to natural speech) and the context in which these sounds were presented (active categorization/discrimination during/after learning versus passive listening), the discrepancy of their results indicates that categorical sound representations depend to a large degree on the context in which categories

are learned and accessed. Evidence from animal electrophysiology suggests that the effect of multisensory context might be modality specific, given that visual signals enter the auditory cortex predominantly via feedback projections from putative multisensory areas in the STS and PFC, whereas the laminar profile from somatosensory stimuli is indicative of forward input in layer 4 (Schroeder & Foxe, 2002). This raises the possibility that learning the same sound categories through audio-visual or audio-tactile training might involve distinct neural processing networks and potentially lead to discriminable category representations.

To test this hypothesis, in the current fMRI study, we employed matched audio-visual (group 1) and audio-tactile (group 2) category learning paradigms to investigate how the qualitatively different but complementary multisensory cues shape the formation of abstract sound categories. We aimed to discern perceptual contrasts between newly learned pitch categories from the responses to a continuous sound space, presented in isolation during passive listening by means of univariate as well as multivariate analysis techniques from different levels along the auditory processing hierarchy.

2 | MATERIAL AND METHODS

Participants. Fourteen volunteers were recruited and reimbursed for their participation in this study. All participants reported normal hearing abilities and had no history of hearing loss or neurological disorder as well as normal or corrected to normal vision. The participants were evenly distributed between two groups, such that half of the participants underwent audio-visual association training (AV, 5 females, mean age: 28 years) and the other half audio-tactile association training (AT, 5 females, mean age: 25 years). All participants gave written informed consent before commencement of any fMRI measurement. The fMRI study was approved by the Ethical Committee Psychology at Maastricht University.

Stimuli. Novel sound categories were constructed from complex artificial ripple sounds (Kowalski et al., 1996a). Ripples were generated using the Matlab (MathWorks) NSL toolbox, available at http://www.isr.umd.edu/Labs/NSL/Software.htm. They varied along three spectro-temporal dimensions, namely fundamental frequency (f_0 in Hz), spectral modulation density (Ω in cycles/octave), and spectral modulation rate (ω *in Hz)* to approximate the rich low-level structure of natural sounds but avoid any pre-existing categorical associations (please refer to **Figure 3.1A** for schematic ripple spectrograms). Ripple density was either 1 or 2 cycles/octave with a fixed modulation depth of 1. The speed of the descending phase drift of the spectral modulation envelope was either 1,2,3,4,5, or 6 Hz. F_0 was varied in a small range of less than one octave according to the category distributions (see below) to adjust ripple pitch. Each sound had a bandwidth of 4 octaves with 1/16 octave frequency spacing. Short (15 ms) linear amplitude ramps were introduced at sound on- and offsets to avoid acoustic clicks caused by sudden energy changes. All sounds had a sampling rate of 44100 Hz and were normalized according to their root mean square values (rms). They were delivered binaurally through MR-compatible in-ear headphones (Sensimetrics).

Visual stimuli consisted of circular patches (diameter 2° visual angle) of square-wave gratings with oblique (45°) black-and-white stripes (spatial frequency 4 cycles/°; onset phase shifted by 1/4 cycle), presented on a uniform grey background. The location of the patches varied between two positions (110° and 160° polar angle) in the upper-left visual field at 5° from the central fixation cross (size: 1°). See **Figure 3.1B** for an example display.

Tactile vibrations were provided through a piezoelectric mechanical stimulation device (Piezostimulator, QuaeroSys Medical Devices, Schotten, Germany). A module consisting of a 2 x 5 pin matrix (pin diameter: 1 mm; center distance between pins: 2.5 mm) was positioned under the tip segment of the little and index finger of the left hand. All ten pins of one module were triggered simultaneously to a height of 0.9 mm with a frequency of ~33 Hz (square wave, timing accuracy: 0.5 ms) for the duration of 1 second. Sport bandages, wrapped around the finger and the

module ensured stable pressure and prevented that modules slipped out of position while subjects rested their hands in a comfortable position on their body. Please see **Figure 3.1C** for a depiction of a stimulation module.

Category distributions. Following our earlier study (Ley et al., 2012), sound categories were defined by ripple pitch (f_0), irrespective of the other spectro-temporal dimensions. For training, six f_0 values were linearly sampled from two non-overlapping normal distributions (standard deviation: 1/8 octave) with category means 1/4 octave from the category boundary (f_0 = 200 Hz). The training sound set thus consisted of 144 different sounds (2 categories with 6 pitches each combined with 6 velocities and 2 densities, for details please refer to **Figure 3.1D**). While sounds close to the category boundary were omitted during training to facilitate the extraction of the relevant feature dimension and formation of two pitch categories, we constructed a second test sound set to assess the categorical nature of perception and sound representations in absence of acoustic categorical cues (i.e. larger f_0 differences at the category boundary relative to within categories). We employed a pitch continuum, consisting of six f_0 values varying in equidistant steps (log space) between category means. The test sound set thus comprised 72 different sounds.

Experimental procedure. Participants started with a pre-training fMRI session followed by three behavioral training sessions and a post-training fMRI session on the last day completed in maximally two weeks. To avoid fading of potential learning effects, behavioral training and post-training fMRI sessions were always conducted on four successive days. The pre-training fMRI session was preceded by a sound familiarization procedure during which participants passively listened to the test sound continuum while lying in the scanner. Behavioral training was conducted inside the magnet (no image acquisition) using the same equipment as during the scanning sessions to match the stimulation context and quality. For an illustration of the procedure, refer to **Figure 3.2**.

Behavioral training. Participants underwent crossmodal association training in which they matched a sound with a concurrent visual (AV training) or tactile (AT training) stimulus appearing at one of two possible locations. Participants maintained fixation throughout the experimental block. In each trial, a randomly chosen sound from one of the two categories was paired with one of the two visual or tactile stimuli. Subjects had to indicate whether the pair was a 'match' or 'no-match' by pressing a button with their right index or middle finger within 3000 ms from stimulus onset. Correct pairings consisted of low-pitch sounds and 110° visual stimulus location or little finger stimulations; high-pitch sounds were correctly paired with 160° visual stimulus location or index finger stimulation. During one training round, each of the 144 training sounds was presented with both visual or tactile stimulus locations, resulting in 288 trials with 50% match and 50% no-match pairings. Participants learned the correct association through trial-and-error, using the provided trial-by-trial visual feedback. Matches were displayed in green, no-matches in red. The subject's response was displayed as a small filled circle (1/2° visual angle) after stimulus offset in the screen center, broad-rimmed (1° visual angle) by the correct response after lapse of the response time (see **Figure 3.3A**). Single-color filled circles thus represented correct responses, while multicolored circles represented a mismatch between the subject's response and the correct response (see **Figure 3.3B**). No information was given with respect to the association rule and participants were instructed to refrain from any verbalization of their matching strategy. To assess the categorical nature of the crossmodal stimulus association after training we tested participant's matching performance without feedback using the 72 test sounds. Analog to the training procedure, each test sound was paired with both stimulus locations, resulting in 144 trials. The results from this generalization test (test-day3) were compared with the same test performed in an early learning stage after one initial feedback-training round (test-day1).

Behavioral data analysis. In the current study, we constrained the analysis of reaction times and accuracies to the two no-feedback generalization tests to

investigate effects of crossmodal many-to-one association training on the categorical nature of pitch perception despite continuous variation of the underlying physical feature (f_0). A more elaborate analysis of behavioral data was reported elsewhere (please refer to **chapter 3** for details). Reaction times were analyzed for correct trials after outlier removal (please consult chapter 3 for details). We inspected the robustness of the crossmodal association within and across sound categories. Thereto, we calculated the differences in percent match responses for both visual/tactile stimuli across the six pitch values irrespective of the other sound dimensions. The resulting data was fitted with a logistic function (see Ley et al., 2012) and slope as well as amplitude differences were tested with a 2 x 2 mixed-design ANOVA using training group (AT and AV) as between-subjects factor and test block (test-day1 and test-day3) as repeated measure. Moreover, reaction times and error rates were compared across the six pitch levels to extract potential boundary effects, typical for Categorical Perception (CP, Harnad, 1987). Error rates and reaction times were averaged for the two boundary pitch levels (3 and 4) as well as for the extremes of the continuum (1, 2, 5, and 6) and the differences were analyzed with a 2 x 2 mixed-design ANOVA with training group as between-subjects factor and pitch range as repeated measures. The behavioral as well as fMRI data from one subject (AV group) were excluded due to insufficient behavioral performance (matching accuracy < 65% on day 3).

Imaging. FMR imaging was performed on a 3-Tesla Siemens Allegra head scanner at the Maastricht Brain Imaging Center (MBIC, Maastricht, the Netherlands). In both pre-training and post-training fMRI sessions, we acquired high-resolution (voxel dimensions = 1 mm isotropic) T1-weighted anatomical images using an ADNI MPRAGE sequence (192 sagittal slices covering the whole brain, repetition time (TR) = 2050 ms; echo time (TE) = 2.6 ms; matrix size = 256 x 256 x 192) for anatomical-functional alignment and cortex reconstruction purposes. In the main experiment, functional T2*-weighted images (TR = 3500 s; acquisition time (TA) = 2400 s; field of view (FoV) = 224 x 224; matrix size = 96 x 96; TE = 30 ms; voxel dimensions =

2.5 mm isotropic) were acquired to measure blood-oxygen-level-dependent (BOLD) contrast while subjects fixated and passively listened to the randomly permuted test sounds. In a clustered volume EPI sequence, 37 slices (10 % slice gap, almost whole brain) were measured in the first 2400 ms of a trial while the sounds were presented during the remaining 1100 ms silent period, leaving a 50 ms gap before and after image acquisition. We employed a slow event-related design (average intertrial interval = 17.5 s, jittered between 4, 5, and 6 TR) for single-trial analysis of response patterns. In both scanning sessions, we acquired three runs of 364 volumes corresponding to 72 test sound trials and four fixation trials at the beginning of each run (discarded from analysis). Stimulus timing was synchronized with MR pulses using Presentation software (Neurobehavioralsystems). Additionally, we employed a functional localizer scan in the post-training session (same imaging parameters as for the main experiment) to identify those regions responding to the ripples, gratings, and vibrotactile stimuli. Here, stimuli of the same condition were presented in six blocks of five TR (one stimulus per TR, presented in the silent gap between acquisitions), alternated with 14 s baseline blocks. Please refer to **chapter 3** for further details on the localizer experiment.

FMRI data analysis. Standard pre-processing of functional data using BrainVoyager QX (Brain Innovations) included slice scan time correction, 3D motion correction and temporal high-pass filtering (please consult chapter 3 for details). Data from one subject (AV group) had to be excluded from the analysis due to excessive head motion (> 3.5 mm). After alignment, both anatomical and functional scans were normalized to Talairach space (Talairach and Tournoux, 1988). Design matrices included the six parameters estimated from the motion correction algorithm to account for motion-related variability in the BOLD signal. Furthermore, we introduced moderate spatial smoothing (FWHM: 4 mm) to the localizer data only. We performed grey- and white matter segmentations to extract individual curvature information used for cortex-based alignment procedures (CBA, Goebel, Esposito, & Formisano, 2006). Functional group analyses were

based on this CBA conversion and the results are displayed on an average mesh reconstruction.

Univariate analysis. To explore learning-induced changes in between-category contrasts, we performed single-subject as well as group (fixed-effects, FFX) analyses of pre- and post-learning fMRI data. For this purpose, we computed a general linear model (GLM) by fitting the BOLD response time course with the predicted time series for low-pitch (levels 1 - 3 of the test-sound continuum) and high-pitch sounds (levels 4 - 6), merging all 36 trials in one category per run, irrespective of velocity and density values. To account for the delay in the hemodynamic response the predicted time courses were convolved with a canonical (double gamma) hemodynamic response function. For each group separately, we tested for interactions between fMRI session and category differences and additionally analyzed the pitch-category contrast in the post-session alone requiring significant above baseline sound responses. To reveal potential pre-existing pitch-category differences, we analyzed the contrast between high-pitch and low-pitch sounds in the pre-session for all subjects together. Here, we restricted the analysis of the localizer data to the responses to vibrotactile stimulation to inspect their overlap with categorical sound representations in the AT group. A more extensive univariate analysis of the localizer and main experiments has been reported previously (see **chapter 3**). Statistical maps were corrected for multiple comparisons on a whole-brain basis using cluster-size thresholding (Forman et al., 1995) yielding a false positive rate of $p < .05$, estimated from 1000 Monte Carlo simulations (initial uncorrected p-values were set to .05 unless indicated otherwise).

Multivariate analysis. The analysis of multi-voxel activation patterns was restricted to the main experiment (auditory only) and constrained by two large anatomical masks (**Figure 4.2A**). One mask covered the entire temporal cortex (TC), encompassing the auditory cortex as well as posterior and medial portions of the lateral sulcus, known to code non-auditory information such as somatosensory stimuli (Schürmann et al., 2006). This

mask was chosen to entail early auditory areas previously shown to represent learning-induced pitch categories (Ley et al., 2012) as well as sites of potential audio-tactile interactions (see **chapter 3**). A second mask covered the posterior parietal cortex (PPC) anteriorly bordered by the postcentral sulcus. This mask included the intraparietal sulcus (IPS), which had displayed overall increased sound responses after AV association learning (see **chapter 3**). The IPS therefore presents a feasible candidate for audio-visual category representations. Both masks were delineated on an average cortex representation and then projected to each individual subject space to achieve a high level of anatomical correspondence between subjects with an account for variability in curvature. The bilateral masks were comparable in size: on average, the temporal lobe mask contained 14110 voxels (*SEM* = 187.7) while the posterior parietal mask contained 12853 voxels (*SEM* = 175.5). With these large masks, we favor a more exploratory approach, which is less driven by specific univariate contrasts that might incorrectly exclude informative voxels due to non-significant average response differences. Furthermore, this approach takes into account that crossmodal learning might engage distributed processes in sensory-specific and multimodal processing areas rather than locally restricted representations.

To allow for a valid comparison across subjects, sessions, masks, as well as trial labeling strategies, we employed fixed parameter settings for trial estimation and classification. In this study, we refrained from any GLM-based voxel pre-selection to circumvent a potential bias towards regions with classical hemodynamic sound responses while neglecting uncommonly shaped and small amplitude responses. Single-trial responses were extracted from the average percent signal change in a time window of two TRs (7s), starting 4.55 s after sound onset relative to one TR preceding sound onset, resulting in 216 response estimates per voxel. We employed linear support vector machines (SVMs) for single-trial classification in combination with a recursive feature elimination procedure (RFE, De Martino et al., 2008). Due to the orthogonal sound design, each trial

response is associated with a pitch as well as a velocity value allowing their division into two equally sized classes according to the trained dimension (i.e. low pitch vs. high pitch) as well as to an untrained dimension (e.g. slow velocity vs. fast velocity), independent of the respective other sound dimensions. The comparison of these two trial labeling strategies allows assessing the specific effect of the behavioral training on the multi-voxel pattern classification in contrast to mere stimulus repetition. Trials were divided into two independent sets for classifier training and testing following a leave-run-out approach, resulting in three different splits. In each split, we employed 100 RFE steps in which two runs were used for classifier training while the remaining run was used to test the classifier's performance and assess the model's generalization ability. After each RFE iteration, the voxels' discriminative weights were ranked and the lowest 5% were discarded. This procedure thus stepwise reduces the number of voxels from ~14000 to ~85 voxels in an attempt to discard noisy and uninformative voxels and retain those voxels that actually drive the between-category separation. Finally, classification accuracies at each RFE level were averaged over the three test data sets to obtain the cross-validation accuracy. The reported values correspond to the highest average accuracy across 100 RFE levels. Due to the iterative voxel elimination procedure and the selection of the maximum accuracy across RFE levels, a chance level of 50% cannot be assumed. Therefore, we conducted a permutation test (Nichols and Holmes, 2002), in which the same RFE procedure was repeated 100 times per subject and session with randomly permuted trial labels to obtain an empirical chance level. The final permutation accuracies per subject reflect the average over 100 best cross-validation accuracies across 100 RFE levels, identical to the real label approach. We tested the learning-induced increase in classification accuracy of pitch categories with a non-parametric one-tailed Wilcoxon matched-pairs signed-ranks test. Furthermore, the same test was employed to investigate whether the post-learning classification accuracy was significantly above the empirically estimated chance level based on permutations.

To identify those voxels, which exhibit a learning-induced increase in their contribution to category discrimination, we computed group rank difference maps. To this end, we determined the 50% most robust voxels per subject and session (i.e. those voxels resisting at least 15 RFE levels in which 50% of the voxels within the mask are eliminated) and ranked them according to their discriminative weight, which reflects their contribution to classification. Subsequently, we computed the post - pre difference to reveal the rank difference of the same voxel's weights between sessions. Single-subject maps were smoothed with a Gaussian kernel (FWHM = 3 mm) to increase their spatial overlap and aligned based on the CBA conversion. Additionally, maps were thresholded with a minimum cluster size of 25 mm^2. To inspect the consistency across subjects, we selected the 25% top ranks per subject in RFE level 15 and computed their overlap.

3 | RESULTS

Behavioral results. The sigmoid curve fitted to the difference in percent match responses between the two visual or tactile stimuli along the relevant dimension (pitch) was characterized by a significant steepening at the category boundary (pitch level = 3.5) with learning ($F_{1,11}$ = 6.46, p = .027). No significant group differences or group interactions could be observed. While this increase in slope reflects better discrimination ability at the category boundary probably due to perceptual learning, the marginally significant increase in amplitude ($F_{1,11}$ = 3.31, p = .096) even for the extremes of the pitch continuum indicates more robust association of the sound categories (high vs. low pitch) with the respective location of the visual or tactile stimulus with training. As can be seen in **Figure 4.1A**, the point of subjective equality (PSE) lied between pitch levels 3 and 4, corresponding to the learned category boundary. The significantly larger reaction times ($F_{1,11}$ = 35.7, p < .001) and error rates ($F_{1,11}$ = 101.43, p < .001) for pitch levels at the category boundary relative to pitch values closer to the category means provide further evidence for the categorical nature of sound perception after training (**Figure 4.1B**). This trend is identical for both

groups, despite overall larger reaction times for the AT training group (M = 1319.2 ms, SD = 88.5 ms) than the AV training group (M = 1003.42 ms, SD = 95.59 ms, $F_{1,11} = 5.88$, $p = .034$).

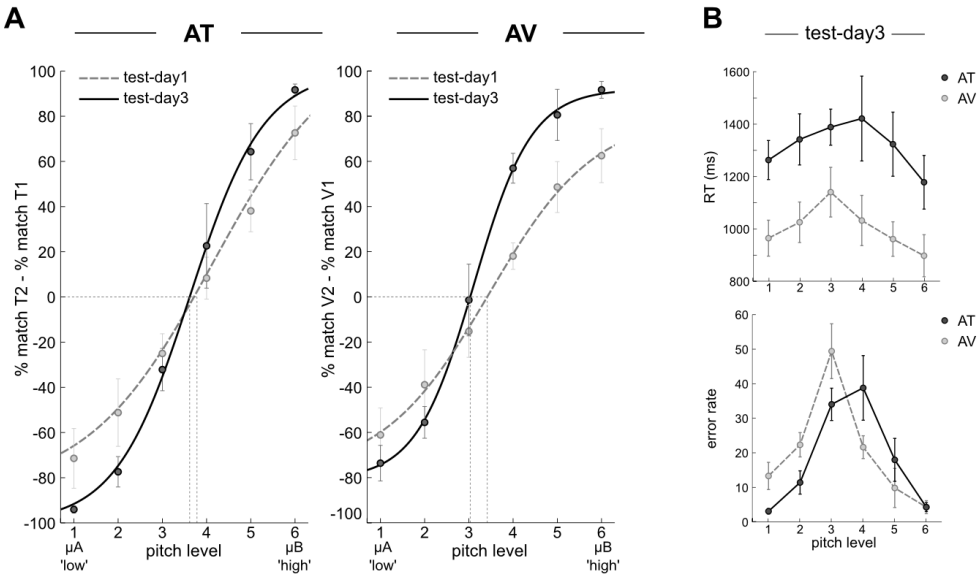


Figure 4.1 Behavioral data during generalization tests (no feedback). A, Difference in percent match performance across pitch levels for the two tactile (left) and visual (right) stimulus positions after 4 feedback training blocks (test-day1) and after additional 12 feedback training blocks (test-day3). T1/V1 correspond to 110° polar angle or little finger, correctly paired during training with low-pitch sounds; T2/V2 correspond to 160° polar angle or index finger, correctly paired during training with high-pitch sounds. Negative values reflect a stronger association with stimulus position 1, positive values a stronger association with position 2. Pitch level 1 corresponds to the category center for low-pitch sounds (μ_l) used for training; pitch level 6 corresponds to the category center for high-pitch sounds (μ_h) respectively. Please note that the test sound continuum consists of novel stimuli not used for training. **B**, Reaction times (top) and error rates (bottom) across the pitch continuum (levels 1 – 6) in the post-learning generalization test (test-day3) after 16 feedback training blocks (1152 feedback trials) for both training groups.

FMRI results - univariate analysis. As expected, significant response differences for the pitch categories could only be detected in a small subset of subjects using contrast-based univariate methods (please refer to **Table**

4.1 for an overview of the single-subject results). Fixed-effects group analyses did not yield any significant learning-related increase in category discrimination. No pre-existing category differences could be revealed prior to learning. Significant post-learning category differences were left lateralized (**Figure 4.2B**). In the AT group, a cluster in the anterior medial Heschl's Gyrus (HG) crossing the first transverse sulcus (FTS) responded significantly stronger to high-pitch than low-pitch sounds after training. Interestingly, the category-selective clusters for the AV group were more spread across the cortex including, additionally to a small cluster in the anterior lateral HG, the superior parietal lobe (SPL) as well as the anterior cingulate cortex (ACC).

	# of subjects	ROI	TAL coordinates (x, y, z)	*p*-value (min cluster size)
Increase in category separation (session x between-category contrast)				
AV	0 out of 5			
AT	1 out of 7	CM	-45, -33, 13 (LH)	$p < .05$ (cs > 21)
Between-category contrast (post-session only)				
AV	1 out of 5	SPL	-13, -68, -51 (LH) 17, -59, 55 (RH)	$p < .05$ (cs > 21)
		lat. HG	49, -19, 10 (RH)	$p < .05$ (cs > 21)
AT	1 out of 7	PIC	46, -33, 23 (RH)	$p < .05$ (cs > 20)
		CM	40, -25, 22 (RH)	$p < .05$ (cs > 20)
		upper bank of STG	50, -25, 12 (RH)	$p < .05$ (cs > 20)

Table 4.1 Single-subject univariate fMRI results. To test for the learning hypothesis, the interaction between fMRI session and pitch-category contrast was defined; additionally, the between-category contrast was tested for the post-training session only. Only significant results ($p < .05$, cluster size (cs) correction) are depicted. CM, caudomedial auditory belt; SPL, superior parietal lobe; lat. HG, lateral Heschl's gyrus; PIC, posterior insular cortex; STG, superior temporal gyrus.

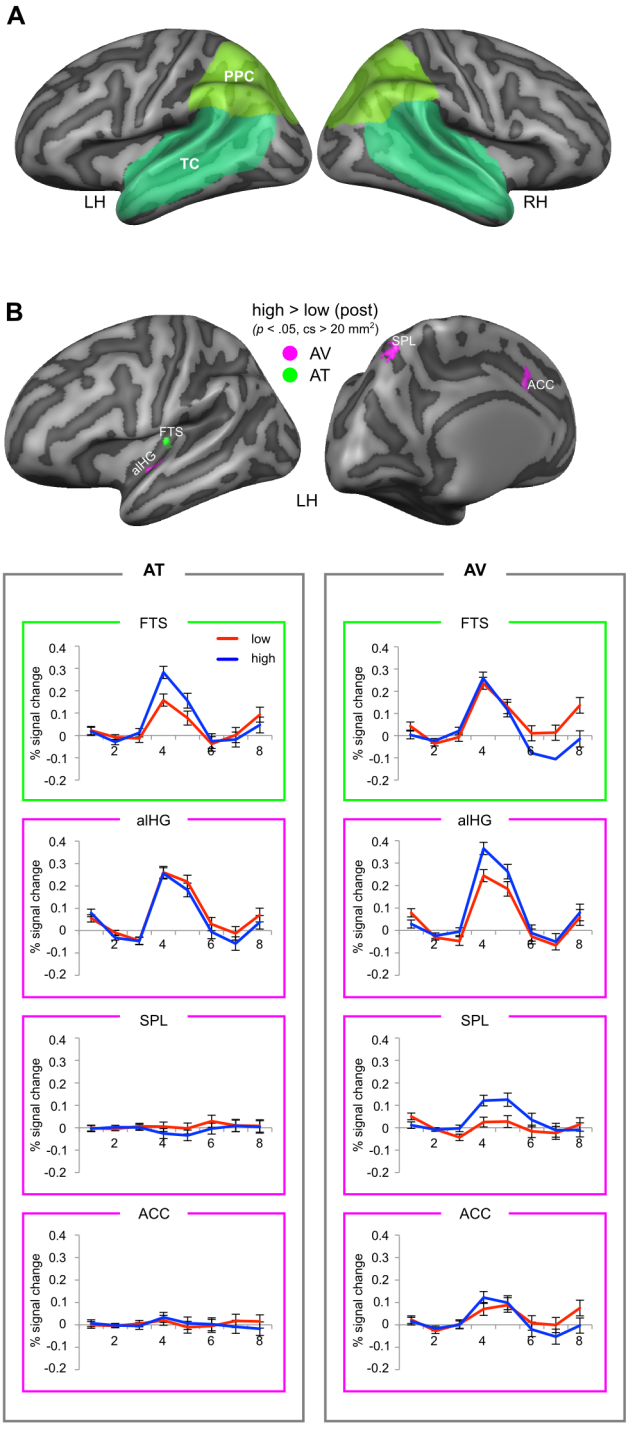


Figure 4.2 Anatomical masks and univariate fMRI results. **A**, Temporal cortex (TC) mask and posterior parietal cortex (PPC) mask depicted on an average cortex reconstruction. **B**, Group (AV = 5, AT = 7, FFX) results for the pitch-category difference (levels 4 - 6 > levels 1 - 3) in the post-training fMRI session. Average time-courses are displayed for the two groups separately (AT, left and AV, right). Frame color indicates for which group the between-category contrast was significant.

FMRI results – multivariate analysis. Abstract pitch categories could successfully be decoded in the AT group only after learning, using the temporal lobe mask. A Wilcoxon matched-pairs signed-rank test indicated that post-learning accuracy (*MDN* = 57.41%) significantly differed from the empirical chance level estimated with permutation (*MDN* = 54.7%, Z = 1.69, p = .046, r = .64). All subjects from the AT group showed an increase in classification accuracy from the pre-training to the post-training session (see **Figure 4.3A**). The median accuracy in the pre-learning session was 53.7%. A Wilcoxon signed-rank test revealed a significant effect of learning (Z = 2.37, p = .009, r = .9). While the median accuracy in the AV group was also 57.41%, the differences to the permutation accuracy in the post-session did not reach significance (p = .25) due to larger inter-subject variance. Furthermore, no learning effect was apparent from pre- to post-training session (p = .173). Classification accuracies for the untrained (i.e. velocity) sound categories did not differ significantly from chance level prior to and also following audio-tactile or audio-visual training. Neither posterior parietal cortex alone nor both masks in combination revealed any successful classification of pitch categories for the AV or AT group. The detailed results for these voxel selections can be viewed in **Table 4.2**.

The maps obtained from the single-subject multivariate pattern analyses confirmed and extended the results from the univariate single-subject and fixed-effects group analyses. The average rank difference maps revealed six main clusters (> 25 mm^2) in the left temporal lobe (**Figure 4.4**). Voxels in early auditory regions on middle HG as well as a more anterior medial region received larger weights during classification in the post versus the pre-training session, indicating increased contribution to the separation of pitch categories. Additionally to two clusters on middle and posterior superior temporal gyrus (STG), a medial posterior region in the auditory belt could be identified, potentially corresponding to macaque area CM (Schroeder et al., 2001; Fu et al., 2003). Another cluster posterior to auditory areas in the caudal end of the lateral sulcus probably corresponds to the posterior insula (PIC) or extensions of the secondary somatosensory

cortex (SII). This cluster overlaps with responses to the vibrotactile stimuli used for audio-tactile training (**Figure 4.4A**). Additional analyses of between-subject overlap revealed that these clusters were consistently present in at least three out of seven subjects.

As can be seen in **Figure 4.4B**, all clusters show a clear average response to the ripple sounds, however, significant between-category differences could only be revealed for the posterior STG cluster after training ($t_6 = 2.83$, $p = .03$). Average signal levels in the middle HG also differed to a considerable degree approaching significance ($t_6 = -2.02$, $p = .09$). It is interesting to note that these two regions show opposite preference for the low- and high-pitch categories.

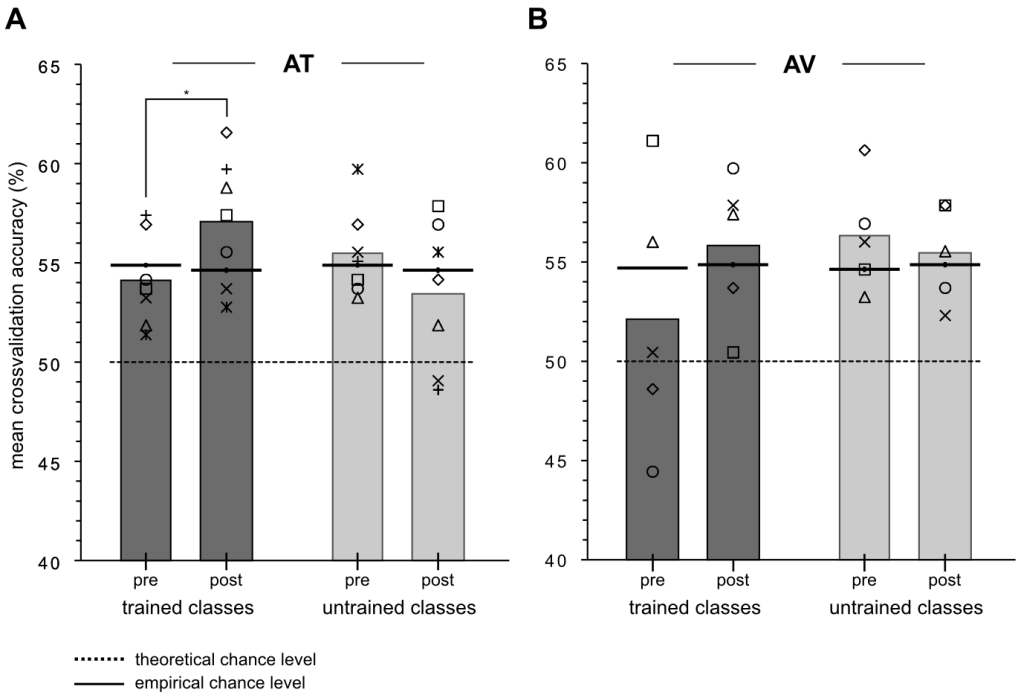


Figure 4.3 Multivariate classification accuracies. Mean (bars) and individual subject (symbols) classification accuracies for the audio-tactile group (**A**) and the audio-visual group (**B**) in the pre-training and post-training fMRI sessions using the temporal cortex mask. Cross-validation accuracies are computed from the average across three splits (leave-run-out). Trained classes correspond to the pitch discrimination (pitch levels 1 - 3 vs. pitch levels 4 – 6); untrained classes correspond to the orthogonal trial separation into slow velocity (ω: 1 - 3 Hz) versus fast velocity (ω: 4 - 6 Hz). The empirical chance level was estimated with 100 trial label permutations for each subject and session separately. Individual subjects are depicted with symbols to trace the classification accuracy across fMRI sessions and trial label procedures; please note that the same symbols are used for AT and AV despite non-overlapping subject groups.

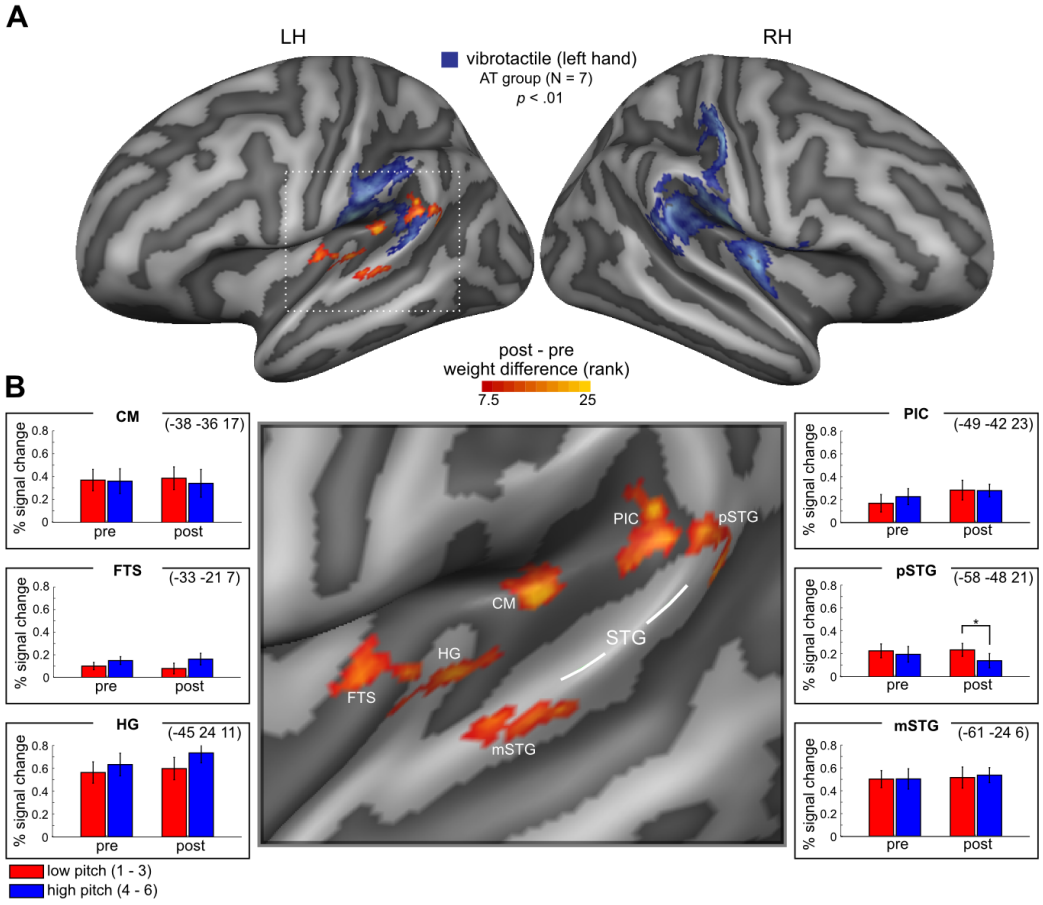


Figure 4.4 Average rank difference maps for the AT group. **A**, Overlap of average rank difference maps with tactile response clusters derived from a separate localizer scan. **B**, Enlarged representation of those voxels that revealed a learning-induced increase in pitch-category discrimination. Bar graphs reflect the average (± *SEM*) signal level across subjects in the 2-TR time window used for classification for the two pitch classes and pre- as well as post-training fMRI sessions. Talairach coordinates (x, y, z) for the main clusters are provided in the individual subplots. * $p < .05$

	PPC				PPC & TC			
	AT		AV		AT		AV	
	pre	post	pre	post	pre	post	pre	post
Pitch (trained)								
MEAN (SEM)	55.95 (1.47)	55.03 (0.99)	55.46 (2.04)	55.00 (1.57)	54.10 (1.38)	54.70 (1.26)	53.24 (1.61)	55.46 (1.57)
MDN	56.02	56.02	53.7	53.7	54.17	54.17	53.24	55.09
Velocity (untrained)								
MEAN (SEM)	55.56 (1.31)	53.51 (0.96)	55.00 (1.03)	57.31 (1.94)	56.08 (0.79)	52.84 (1.32)	54.72 (1.65)	56.39 (1.44)
MDN	55.09	53.24	55.09	56.02	54.63	53.24	54.63	56.94
Permutation (chance)								
MEAN (SEM)	54.10 (1.38)	54.70 (1.26)	53.24 (1.61)	55.46 (1.57)	54.64 (0.13)	54.44 (0.07)	54.24 (0.11)	54.73 (0.18)
MDN	54.17	54.17	53.24	55.09	54.45	54.50	54.16	54.66

Table 4.2 Average classification accuracies for the parietal and temporal-parietal masks. Classification accuracies (mean cross-validation accuracy across splits) for the posterior parietal cortex (PPC) mask and the combined masks encompassing the temporal as well as posterior parietal cortex (PPC & TC) for the audio-tactile (AT) and audio-visual (AV) groups.

4 | DISCUSSION

In this fMRI study we aimed to investigate the influence of crossmodal interactions during association training on the formation of categorical sound representations. We analyzed neural response patterns elicited by passive listening to complex ripple sounds using MVPA before and after audio-visual or audio-tactile association training to reveal potential differences in learning-induced categorical representations of the same acoustic input. The behavioral results show that the many-to-one crossmodal association training increased the perceptual separation of ripple pitch at the trained category boundary. This effect was present for both audio-visual and audio-tactile learners despite large within-category feature variability. However, only after audio-tactile training, abstract pitch categories could successfully be decoded from distributed activation patterns in the temporal lobe. Relevant temporal regions included early auditory regions and audio-tactile integration areas in the posterior end of the lateral sulcus.

No pre-existing representation of abstract pitch categories

Prior to crossmodal training, neither univariate group analyses nor single-subject multivariate analyses revealed a significant discrimination between high-pitch and low-pitch categories. This also holds for an orthogonal separation of the stimulus space into sounds with slow versus fast modulation rates. This result can be explained by the small range and gradual variation of the low-level feature dimensions crossing the defined category boundary in the test sound space, which does not lead to a categorical percept per se. This is confirmed by the shallow slope of the average crossmodal association curve after one initial training round. Furthermore, the spectro-temporal profile of the sounds within categories was not homogenous but differed along two irrelevant feature dimensions posing an *invariance problem* (Myers et al., 2009). During training, 36 different low-pitch sounds were associated with the tactile stimulation of the left little finger or a visual stimulus at 110° polar angle and 36 different high-

pitch sounds with the same stimulus at the left index finger or 160° stimulus position. This association task required the extraction of the sound properties relevant for the correct crossmodal association (i.e. pitch) and the formation of abstract object categories invariant to irrelevant temporal and spectral sound modulations. Apart from an overall slower response rate for audio-tactile associations, both groups performed equally well and their crossmodal matching behavior suggests the formation of a categorical pitch percept. This is reflected in their s-shaped crossmodal association curves, which resemble identification curves for phoneme categories (Liberman et al., 1957).

Formation of category representations in early auditory cortex

The analysis of post-learning BOLD responses to the pitch continuum revealed that audio-tactile learning led to a training-specific increase in pitch-category discrimination in multivoxel patterns in the temporal lobe. In contrast, the classification performance for the untrained velocity categories was characterized by a minor but non-significant decrease with learning. This suggests that audio-tactile association training promoted the extraction of the diagnostic sound feature and changed the perceptual representation of the ripple sounds leading to an increased sensitivity for the between-category contrast at early auditory processing levels. The involvement of early auditory areas on STG and HG is in line with our previous findings obtained with explicit unisensory category training (Ley et al., 2012) and provides further evidence for categorical representations at sensory processing levels. This result corroborates recent findings of abstract sound processing functions in the posterior auditory cortex for non-speech categories (Giordano et al., 2012) and contradicts the prevalent view that categorical sound processing generally relies on neural resources in classical speech areas in the left STS (Leech et al., 2009; Liebenthal et al., 2010). The recent addition of advanced fMRI analysis techniques (Lee et al., 2012) and the extension of the commonly used stimulus material from speech or speech-like sounds to other natural categories (e.g. musical

chords, Klein & Zatorre, 2011) as well as the use of learning paradigms (Myers & Swan, 2012) revealed that categorical sound processing is rather distributed over bilateral cortex, ranging from low-level feature analysis stages to higher levels of decision and motor processes in the frontal lobe (Lee et al., 2012; Myers & Swan, 2012). This diversity bears evidence for the variety of processes underlying sound categorization that depend to a large degree on the employed stimulus material and task.

Crossmodal contribution

Notably, in addition to early auditory cortex, category relevant representations included the posterior auditory belt. This area has consistently been found in electrophysiology and fMRI studies in response to purely tactile stimulation in animals (Schroeder et al., 2001; Fu et al., 2003; Kayser et al., 2005) and humans (Foxe et al., 2002; Schürmann et al., 2006; Nordmark et al., 2012). While our localizer data did not reveal any somatosensory activation in the same area with the employed threshold, we found an overlap of tactile responses with sound category representations in the posterior end of the lateral sulcus, corresponding to the posterior pole of the insula or extensions of the secondary somatosensory regions (SII). This functional correspondence suggests that crossmodal interactions during learning exerted influence on the formation of categorical pitch representation. Due to the close feature correspondence of auditory and vibrotactile stimuli, association learning might have formed crossmodal representations that facilitate the read-out of abstract pitch information through complementary auditory and somatosensory processing. This interpretation accords with earlier observations of increased sound sensitivity during concurrent vibrotactile stimulation (Schürmann, Caetano, Jousmäki, & Hari, 2004) as well as frequency-specific effects of tactile distractors on auditory pitch-discrimination (Yau et al., 2010). Specifically, Igushi et al. (2007) reported that co-activation of auditory cortex and SII significantly improved vibrotactile frequency discrimination, even when the auditory stimulus did not coincide with the vibrotactile stimulus. These

powerful crossmodal interactions probably originate from direct anatomical connections between auditory and somatosensory regions (Ro et al., 2012) and via the posterior insular cortex as an important polymodal hub for crossmodal information exchange (Hackett et al., 2007). Hence, it is plausible that listening to the isolated sounds, which have been paired with vibrotactile stimuli during training, engaged tactile in addition to auditory processes to enhance frequency sensitivity. As subjects were not engaged in an active association task, it is less likely that this crossmodal activation is caused by imagery of the vibrotactile component, although this explanation cannot be excluded in the current experimental context.

Left hemisphere specialization?

Our findings show that the most discriminative voxels for the newly learned pitch categories are left lateralized. One possible but non-exclusive explanation for these left-lateralized effects might be a functional specialization of left auditory cortex for audio-tactile interactions. A left-hemispheric dominance for somatosensory input in the auditory cortex has been reported before (Gobbelé et al., 2003; Schürmann et al., 2006). Yet, we believe that another plausible contribution to the observed left-lateralized effects might be larger inter-individual anatomical variability of the right superior temporal cortex (STC) (Van Essen, 2005), which would exert a strong influence on the alignment of individual curvature and potentially abolish functional activation clusters on the group level. Recently, cross-sectional studies have shown that the right auditory cortex is more prone to developmental changes than the left, leading to prominent anatomical and functional asymmetries in adults (Bonte et al., 2013; Sowell et al., 2002). This age-related increase in inter-subject variability of the right STC might be caused by experience-dependent differences in individual development in contrast to consistent language-related functional development of the left hemisphere (Im et al., 2010). In our subject sample, the inspection of the sulcal morphology in the STC (determined in the anatomical scans for each individual in volume space) revealed a larger

variability in the right hemisphere (4 subjects with common stem HG duplications; 1 subject with a complete HG duplication) compared to the left hemisphere (2 subjects with common stem HG duplications) due to complete or partial HG duplications. These hemispheric differences in macroanatomy are in agreement with earlier reports (Leonard, 1998) and might exert a large influence on the CBA results.

Lastly, left auditory cortex might possess a processing advantage in the employed task due to its higher temporal resolution (Zatorre & Belin, 2001; Schönwiesner et al., 2005) that might help to inhibit processing of the prominent but irrelevant time-varying features in the ripple sounds. The complex spectro-temporal structure of the ripple sounds certainly required the suppression of velocity and density differences to perceptually minimize within-category variability. However, this last option is considered rather unlikely, for the categorization of the ripples according to their abstract pitch values required complex spectral sound analysis, a function commonly attributed to the right auditory cortex (Zatorre & Belin, 2001; Okamoto et al., 2009).

Audio-tactile vs. audio-visual

In contrast to the AT group, classification accuracies after audio-visual training did not exceed the empirical chance level, despite an increase in accuracy from pre- to post-learning in four out of five subjects for the temporal lobe mask. Due to the necessity to exclude two from the originally seven participants, it is possible that these non-significant results are due to insufficient sensitivity. There are however additional factors, which might explain this outcome. First of all, it can be assumed that a crossmodal processing strategy for pitch discrimination is more likely for audio-tactile than audio-visual learning due to the similarity in frequency-based processing mechanisms (Soto-Faraco & Deco, 2009). Despite comparable behavioral performance, the crossmodal matching might therefore require more effort and become less automatic for audio-visual than audio-tactile stimulus pairs and consequently involve different levels of crossmodal

interaction. While the audio-tactile interface may be the auditory cortex and neighboring multimodal areas, audio-visual category learning may require the involvement of attentional mechanisms in higher-order association cortex. This interpretation is in line with the training-specific enhancement of sound responses in audio-visual spatial attention networks in the PPC for the AV group, whereas AT training effects were restricted to the temporal lobe (see **Chapter 3**). Our attempt to include these higher-order processing regions in the analysis by using posterior parietal as well as combined temporal-parietal masks did not yield any significant results. This can be ascribed to the passive paradigm chosen in the current experiment, which may be suboptimal for the engagement of fronto-parietal networks. Former studies have shown that the parietal network for visuo-spatial attention is particularly involved during early phases of active category learning (Little & Thulborn, 2005) and that frontal activation levels significantly decrease with the level of expertise (Myers et al., 2009). It would require an active categorization task to definitely resolve this matter. Furthermore, the large number of voxels in the combined masks may have lowered the signal-to-noise ratio (SNR) significantly; a factor known to impair classifier performance (Norman et al., 2006).

Source of improved classification accuracy

Our highly controlled stimulus design allows to exclude stimulus-related as well as familiarity or repetition-related effects as source of increased classification accuracy in the post-training fMRI session for the AT group. The specific increase in classifier performance for the trained pitch classes clearly suggests a learning-induced effect on the multivoxel sound representations. It is noteworthy that the overall signal level in the temporal lobe mask in the 2-TR time window used for classification did not significantly differ between the pre- and post-learning fMRI sessions using the same acoustic input, which further eliminates global amplitude differences as driving factor (Smith et al., 2011). It is difficult to determine whether the observed effect is truly multivariate (i.e. caused by spatially

distributed response patterns rather than pooling of local response differences). The inspection of the signal level in the average rank difference maps and their overlap with the univariate fixed-effects group results points towards a combination of both. Early areas on HG and FTS seem to comprise category selective neural populations overlapping with areas of high-frequency preference in tonotopic maps (Formisano et al., 2003; Moerel et al., 2012). In contrast, the average signal in more posterior clusters covering secondary auditory cortex is mostly non-indicative of pitch category pointing towards a more distributed effect.

The results presented here reveal intriguing differences between audio-visual and audio-tactile learning, suggesting that only the latter promotes sustainable representations of abstract pitch information by means of low-level crossmodal integration of stimulus information. These findings provide important evidence for the eminent crossmodal influence that shapes *where* and *how* the same sounds are represented. Hence, combining MVPA techniques with multisensory learning paradigms opens up new possibilities to study crossmodal neural plasticity.

REFERENCES

Bonte M, Frost MA, Rutten S, Ley A, Formisano E, Goebel R (2013) Development from childhood to adulthood increases morphological and functional inter-individual variability in right superior temporal cortex. Neuroimage 83: 739-750.

Van Essen DC (2005) A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. NeuroImage 28: 635–662.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 33: 636–647.

Formisano E, Kim DS, Di Salle F, Van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40: 859–869.

Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM (2002) Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J Neurophysiol 88: 540–543.

Fu K-MG, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE (2003) Auditory cortical neurons respond to somatosensory stimulation. J Neurosci 23: 7510–7515.

Giordano BL, McAdams S, Zatorre RJ, Kriegeskorte N, Belin P (2012) Abstract encoding of auditory objects in cortical activity patterns. Cereb Cortex: 1–13.

Gobbelé R, Schürmann M, Forss N, Juottonen K, Buchner H, Hari R (2003) Activation of the human posterior parietal and temporoparietal cortices during audiotactile interaction. NeuroImage 20: 503–511.

Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. Hum Brain Mapp 27: 392–401.

Hackett TA, Smiley JF, Ulbert I, Karmos G, Lakatos P, De la Mothe LA, Schroeder CE (2007) Sources of somatosensory input to the caudal belt areas of auditory cortex. Perception 36: 1419–1430.

Harnad S (1987) Categorical Perception: The Groundwork of Cognition (Harnad S, ed). Cambridge, UK: Cambridge UP.

Husain FT, Fromm SJ, Pursley RH, Hosey L, Braun A, Horwitz B (2006) Neural bases of categorization of simple speech and nonspeech sounds. Hum Brain Mapp 27: 636–651.

Iguchi Y, Hoshi Y, Nemoto M, Taira M, Hashimoto I (2007) Co-activation of the secondary somatosensory and auditory cortices facilitates frequency discrimination of vibrotactile stimuli. Neuroscience 148: 461–472.

Im K, Jo HJ, Mangin J-F, Evans AC, Kim SI, Lee J-M (2010) Spatial distribution of deep sulcal landmarks and hemispherical asymmetry on the cortical surface. Cereb Cortex 20: 602–611.

Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. Neuron 48: 373–384.

Kayser C, Petkov CI, Logothetis NK (2009) Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. Hear Res 258: 80–88.

Klein ME, Zatorre RJ (2011) A role for the right superior temporal sulcus in categorical perception of musical chords. Neuropsychologia 49: 878–887.

Kowalski N, Depireux DA, Shamma SA (1996) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J Neurophysiol 76: 3503–3523.

Lee Y-S, Turkeltaub P, Granger R, Raizada RDS (2012) Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. J Neurosci 32: 3942–3948.

Leech R, Holt LL, Devlin JT, Dick F (2009) Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. J Neurosci 29: 5234–5239.

Leonard C (1998) Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? Cereb Cortex 8: 397–406.

Ley A, Vroomen J, Hausfeld L, Valente G, De Weerd P, Formisano E (2012) Learning of new sound categories shapes neural response patterns in human auditory cortex. J Neurosci 32: 13273–13280.

Liberman AM, Harris KS, Hoffman HS, Griffith BC (1957) The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol 54: 358–368.

Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR (2010) Specialization along the left superior temporal sulcus for auditory categorization. Cereb Cortex 20: 2958–2970.

Little DM, Thulborn KR (2005) Correlations of cortical activation and behavior during the application of newly learned categories. Cogn Brain Res 25: 33–47.

De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. NeuroImage 43: 44–58.

Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J Neurosci 32: 14205–14216.

Myers EB, Blumstein SE, Walsh E, Eliassen J (2009) Inferior frontal regions underlie the perception of phonetic category invariance. Psychol Sci 20: 895–903.

Myers EB, Swan K (2012) Effects of category learning on neural sensitivity to non-native phonetic categories. J Cogn Neurosci 24: 1695–1708.

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15: 1–25.

Nordmark PF, Pruszynski JA, Johansson RS (2012) BOLD responses to tactile stimuli in visual and auditory cortex depend on the frequency fontent of stimulation. J Cogn Neurosci 24: 2120–2134.

Norman K a, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10 :424–430.

Ohl FW, Scheich H, Freeman WJ (2001) Change in pattern of ongoing cortical activity with auditory category learning. Nature 412: 733–736.

Okamoto H, Stracke H, Draganova R, Pantev C (2009) Hemispheric asymmetry of auditory evoked fields elicited by spectral versus temporal stimulus change. Cereb Cortex 19: 2290–2297.

Ro T, Ellmore TM, Beauchamp MS (2012) A neural link between feeling and hearing. Cereb Cortex: 1–7.

Russ BE, Lee Y-S, Cohen YE (2007) Neural and behavioral correlates of auditory categorization. Hear Res 229: 204–212.

Schönwiesner M, Rübsamen R, Von Cramon DY (2005) Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur J Neurosci 22: 1521–1528.

Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cogn Brain Res 14: 187–198.

Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC (2001) Somatosensory input to auditory association cortex in the macaque monkey. J Neurophysiol 85: 1322–1327.

Schürmann M, Caetano G, Hlushchuk Y, Jousmäki V, Hari R (2006) Touch activates human auditory cortex. NeuroImage 30: 1325–1331.

Schürmann M, Caetano G, Jousmäki V, Hari R (2004) Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. J Acoust Soc Am 115: 830.

Smith AT, Kosillo P, Williams AL (2011) The confounding effect of response amplitude on MVPA performance measures. NeuroImage 56: 525–530.

Soto-Faraco S, Deco G (2009) Multisensory contributions to the perception of vibrotactile events. Behav Brain Res 196: 145–154.

Sowell ER, Thompson PM, Rex D, Kornsand D, Tessner KD, Jernigan TL, Toga AW (2002) Mapping sulcal pattern asymmetry and local cortical surface gray matter distribution in vivo: maturation in perisylvian cortices. Cereb Cortex 12: 17–26.

Stein BE, Meredith MA (1993) The Merging of the Senses. Cambridge, Massachusetts: MIT Press.

Talairach J, Tournoux P (1988) Co-Planar Stereotaxic Atlas of the Human Brain. Stuttgart: Thieme.

Yau JM, Weber AI, Bensmaia SJ (2010) Separate mechanisms for audio-tactile pitch and loudness interactions. Front Psychol 1: 1–11.

Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11: 946–953.

Chapter 5

GENERAL DISCUSSION

Experience is not what happens to a man;
it is what a man does with what happens to him

~ Aldous Huxley

The concept of brain plasticity, i.e. the susceptibility of the neural system to changes of behavior, experiences, and pathology, dates back to the early 20th century. Even when its mechanisms at the neuronal level were far from being explored at that time, the existence of neuroplasticity was manifested in development and as a consequence of injuries. More recently, it has been accepted that even low-level sensory cortices remain malleable throughout lifespan, and although to a limited extent, change as a function of learning (for reviews, see Pascual-Leone, Amedi, Fregni, & Merabet, 2005; Weinberger, 1995). Plasticity is the fundamental mechanism underlying the remodeling of the neural circuits according to the environmental demands, which is the basis for efficient and goal-directed behavior. During the learning of new categories, plasticity takes effect in selectively enhancing the processing of behaviorally relevant information at the expense of details that do not serve the discrimination between categories. It is assumed that successful category learning results in a reduced and more abstract perceptual representation of the same sensory input, which optimizes the information read-out for task-related purposes.

1 | THE CONTRIBUTION OF THIS THESIS

Quasi-permanent changes of the human auditory system as a consequence of maturation as well as sensory loss (Kuhl et al., 1992; Levänen et al., 1998; for review, see Kral et al., 2002) have been studied quite extensively in the past. In contrast, rapid neural reorganization of sound representations as a function of learning new categories remains largely unexplored. Especially the role of lower-level auditory areas in categorical processing is still under debate. Furthermore, despite considerable evidence for non-acoustic input in the human auditory cortex (Calvert et al., 1997; Foxe et al., 2002; Pekkola et al., 2005; Schürmann et al., 2006; Nordmark et al., 2012), knowledge on the effect of multisensory interaction on learning-induced plasticity is limited.

This thesis aimed to gain more insights into the formation of behaviorally relevant sound representations with a special emphasis on the presence of

rapid learning-induced plasticity in the human auditory cortex. For this purpose, the empirical imaging studies presented here compared the neural responses to artificial ripple sounds before and after successful learning to group them into distinct pitch categories. In addition to the conventional univariate analysis of learning-induced changes in sound responses, **chapter 2** and **4** employed multivariate analysis techniques to decipher perceptual representations of newly learned sound categories from spatially distributed brain activation patterns.

Secondly, this thesis addressed the role of multisensory contextual influence during category learning on subsequent sound processing. Specifically, **chapter 3** and **4** provided a unique comparison of audio-visual and audio-tactile training paradigms to elucidate potential differences in crossmodal plasticity. Thereto, both studies employed a complex many-to-one crossmodal spatial association task. Here, sound categories are learned without explicit labels but through the co-occurrence of the sounds from one category with a tactile or visual cue at one of two distinct spatial locations.

Together, the findings from these three studies help to shed light onto the natural but non-trivial capacity of the auditory system to process the enormous diversity of sounds in an adaptive, goal-directed fashion.

2 | LEARNING AS OPTIMAL METHOD TO STUDY ABSTRACTION AND PLASTICITY OF SOUND REPRESENTATIONS

Humans recognize the voice of a familiar person with ease, even under highly variable conditions in which the signal reaching the ears is a mixture of various voice-specific and unspecific acoustic cues. Over time, they learn to separate relevant from irrelevant acoustic information by integrating sensory and experience-related information. This results in enhanced perceptual discrimination despite unchanged physical input. These changes in sound perception imply changes of the neural sound representation from an acoustic mapping of the sound features to a perceptual representation that selectively emphasizes the characteristic voice features. In order to gain

insights into this experience-related neural transformation, it is insufficient to study the representation of overlearned familiar sound categories such as speech, voices, or music, but rather requires controlled learning paradigms of new sound categories. With novel, artificial sounds, the acoustic properties can be controlled, such that physical and perceptual representations can be decoupled and interference with existing category representations can be avoided. A comparison of pre- and post-learning sound representations provides information about the abstraction capacity at different levels of the sound processing hierarchy. At the same time, it provides insights into the amenability of sound representations to learning-induced plasticity.

3 | RAPID LEARNING-INDUCED PLASTICITY IN EARLY AUDITORY CORTEX

The experiment presented in **chapter 2** was mainly motivated by results from animal electrophysiology. Recordings in the gerbil primary auditory cortex (A1) had revealed a learning-induced transformation from a frequency-specific coding to a categorical (upward vs. downward sweep) coding of the same acoustic stimuli (Ohl et al., 2001). Crucially, between-category differences were expressed in spatiotemporal firing patterns rather than overall activation level. These results have interesting implications for hierarchical models of auditory processing in which A1 is considered as site of basic feature extraction and complex (including categorical) processes are accomplished only with the aid of progressively higher processing stages in the anterior temporal or prefrontal cortex (PFC) (Rauschecker & Scott, 2009). Psychophysical learning experiments in humans have provided strong evidence for perceptual transformations induced by category learning. These transformations are typically expressed in perceptually enhanced between-category differences relative to within-category differences despite equal physical distances (Guenther et al., 1999; Goudbeek et al., 2009). However, the underlying changes in the neural sound representations remain largely elusive. The lack of evidence in this respect can be attributed

to the fact that the between-category contrast is a purely perceptual phenomenon. On top, members of a single category can differ along several irrelevant dimensions. Disentangling perceptual and acoustic representations thus poses great difficulties for conventional univariate analysis techniques that rely on changes in overall activation level.

Chapter 2 demonstrated that multi-voxel pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) data could reveal a previously unexplored reorganization of sound representations also in early human auditory cortex as a consequence of category learning. The results show that after successful learning, locally distributed response patterns in Heschl's gyrus (HG) and its adjacency become selective for the trained category discrimination despite unchanged physical input. Critically, the similarity of the cortical activation patterns correlated with perceptual rather than physical sound similarity, further supporting the role of the early auditory cortex in perceptual object representation rather than acoustic feature mapping (Nelken, 2004). It is noteworthy that these abstract categorical representations were detectable despite passive listening conditions. This is an important detail, as it demonstrates that categorical representations are (at least partially) independent of higher-order decision or motor-related processes. Furthermore, it suggests that some preparatory (i.e. multipurpose) abstraction of the physical input happens at the level of the auditory cortex. How these abstraction processes are implemented at the neuronal level cannot be assessed with the methods employed in the context of this thesis. However, it is plausible that three days of training sufficed to reshape the spectral tuning properties of auditory neurons according to the relevant feature dimension and range. The large spectro-temporal variability within pitch categories further suggests that the response profiles of the neuronal populations become invariant to irrelevant fluctuations in the signal, building the basis for perceptual constancy within categories. The s-shape of the behavioral category identification curve might be an indication for the tuning width along the relevant dimension. While the tuning might be rather coarse within category centers (i.e. around 168 Hz and 238 Hz) to tolerate irrelevant variability, a sharper discrimination is

required at the category boundary (i.e. at 200 Hz) to detect a change in category identity. Recent evidence suggests that spectro-temporal tuning profiles are prone to context modulations (Fritz et al., 2003; David et al., 2012), however, these category-specific effects are far from being understood.

Nevertheless, **chapter 2** makes a strong contribution to the ongoing debate about the level of the auditory processing hierarchy at which abstract categorical representations are implemented in humans. It highlights the special role of early auditory areas in the low-level abstraction of sound representations for categorization.

4 | LEARNING-INDUCED CROSSMODAL PLASTICITY AT DIFFERENT PROCESSING LEVELS

The findings of chapter 2 revealed the flexibility of sound representations in the auditory cortex, making them amenable to changes in environmental demands. As the perception of natural environments is essentially multisensory, it is likely that sound representations are also prone to influences from the visual or tactile modality. Support for this claim comes from a number of studies showing response modulation of auditory neurons by simultaneous non-acoustic events and even crossmodal activation of the auditory cortex in absence of sound stimuli (Schroeder & Foxe, 2002; Fu et al., 2003; Brosch et al., 2005; Kayser et al., 2005; Schürmann et al., 2006) challenging the traditional hierarchical model of multisensory integration in higher-order polymodal processing sites. The early convergence of auditory, visual, and tactile information in the auditory cortex enables crossmodal modulations of putative low-level auditory processes such as the learning-induced reorganization described in chapter 2. This observation motivated the research question of **chapter 3**, namely: How do multisensory experiences during learning influence subsequent sound processing? In particular, does short-term multisensory association training result in crossmodal reorganization of the sound representation? Observations from natural life-long multimodal experiences such as voices and faces suggest

that the functional coupling of the representations from the two modalities results in crossmodal plasticity, such that voices alone activate face-processing areas and vice versa (Calvert et al., 1997; Pekkola et al., 2005; von Kriegstein et al., 2005). In this respect, it is important to note that animal electrophysiology has revealed different laminar profiles for tactile and visual signals in the auditory cortex indicative for forward and backward directed input respectively (Schroeder & Foxe, 2002). This suggests that the influence of visual and tactile stimuli on subsequent sound processing might be mediated via different processing networks. Thus far, little attention has been paid to the differences in crossmodal processing for audio-visual and audio-tactile stimulus pairs.

The study presented in **chapter 3** revealed that three days of multisensory association training suffice to induce a rapid task-related reorganization of the sound representations from an acoustic feature mapping to a multimodal representation at different processing levels, depending on the involved modality. The comparison of matched audio-visual and audio-tactile association training paradigms revealed a prominent discrepancy between post-learning sound responses. Audio-visual association training resulted in a remapping of sound processing onto higher-order visual areas in the fronto-parietal cortex, which did not respond to the same sounds prior to learning. Conversely, audio-tactile association training led to an increase in sound responses in auditory cortex and posteriorly adjacent areas involved in somatosensory processing. These results suggest that crossmodal plasticity can be induced for newly learned arbitrary bimodal stimulus pairs in a short time and confirm the earlier assumption that multisensory interaction during learning affects *how* sounds are processed subsequently in isolation. The dissociation between 'low'- and 'high'-level remapping for audio-tactile and audio-visual learning respectively indicates that crossmodal association learning was mediated by distinct processing networks. This is in line with anatomical studies, which suggest that the source of visual input in the auditory belt lies in multisensory integration areas in the fronto-parietal lobe, whereas tactile input in the auditory cortex may be additionally

mediated via direct connections between auditory and somatosensory regions (Hackett et al., 2007; Ro et al., 2012).

5 | INFLUENCE OF MULTISENSORY CONTEXT ON CATEGORICAL SOUND REPRESENTATIONS

The striking effect of crossmodal plasticity observed in chapter 3 raised the question whether the new sound representations formed through crossmodal association learning also play a functionally relevant role in pitch categorization. Furthermore, the discrepancy between audio-visual and audio-tactile training effects suggests that categorization of the same sounds might be mediated by different processing networks. The rationale of **chapter 4** was therefore to explore changes in categorical sound representations induced by crossmodal association training and unravel potential differences in the neural representations of the same sound categories for audio-visual versus audio-tactile conditions.

The results revealed that audio-tactile association learning-induced the formation of category-selective activation patterns in the auditory cortex and audio-tactile integration areas in the posterior end of the lateral sulcus. This further supports the role of the auditory cortex in abstract sound processing, and additionally indicates that crossmodal interactions during learning influenced the formation of categorical pitch representations. Despite the lack of bottom-up vibrotactile stimulation during scanning, the sounds alone seem to engage tactile processes in the posterior insula. The facilitative effect of co-activation of auditory and tactile processing regions has been demonstrated before for complementary auditory and tactile stimuli (Iguchi et al., 2007) even during asynchronous stimulus presentation, indicative of a task-dependent crossmodal processing strategy. It is proposed that the intimate feature correspondence of auditory and vibrotactile stimuli may have established meaningful crossmodal links between the auditory and tactile representations during association training, which are engaged during subsequent categorical sound processing.

The results from audio-visual training were less conclusive and did not reveal category-selective response patterns in the auditory cortex. While insufficient sensitivity cannot be excluded with certainty as potential source of these results, chapter 3 already indicated that the primary locus of audio-visual plasticity lied in high-level fronto-parietal rather than low-level auditory areas. Frontal areas are specifically recruited during early phases of active category learning involving attention and decision-related processes (Boettiger & D'Esposito, 2005; Little & Thulborn, 2005; Li et al., 2009). Thus, the passive design employed in the current study may have been suboptimal to reveal putative categorical sound representations in fronto-parietal cortex. Furthermore, it should be considered that audio-visual integration through multisensory relay stations might require prolonged training to shape low-level representations in the auditory cortex via feedback projections. In contrast, low-level integration of auditory and tactile stimuli may exhibit an immediate effect on auditory neurons.

6 | OVERALL CONCLUSIONS AND IMPLICATIONS

In summary, the results from the three studies presented in this thesis provide a marked demonstration for the dynamics of auditory processing. In line with the dynamicist view (Engel, 2001), the findings suggest that neural sound representations are not static but continuously updated by contextual information (also from other senses) and prior knowledge to build functionally relevant representations that could correctly guide meaningful behavior. The learning studies showed that repeated exposure to certain contextual conditions induces sustained changes in the sound representations, which can be observed even past the scope of the learning environment during passive listening to the sounds in isolation. Furthermore, this research corroborates the emerging view that the human auditory cortex mediates complex perceptual processing operations beyond basic feature analysis. The successful decoding of perceptual categories from early auditory cortex after category learning suggests that the activation of distributed neuronal populations provides a low-level abstraction of

acoustic features, building the basis for categorical perception. The selective enhancement of the behaviorally relevant features through rapid task-related plasticity of the neuronal tuning properties may be sufficient to achieve a reduced multipurpose representation at these lower-level stages that could be propagated to the frontal cortex to be transformed into a motor code if the task requires it. In this way, a progressively higher abstraction level is achieved in an efficient manner. Specific modules for ecologically valid and highly familiar sound categories such as native speech phonemes or voices (Belin et al., 2000; Leaver and Rauschecker, 2010) might have emerged through long-term adaptation. Constantly recurring stimulation could lead to (mostly) irreversible modifications of the spectro-temporal response profiles of neurons to optimally adjust the processing needs. Speech-like sounds with similar acoustic properties would engage these 'speech-specific' regions to exploit their feature sensitivities (Desai et al., 2008; Leech et al., 2009).

As the focus of the current studies lied on the comparison of the sound representations prior to and following learning rather than the learning process itself, the interactions between frontal and temporal regions mediating the transformation of low-level sound representations remain unexplored at this point. It is likely that task-related processes in the frontal lobe shaped the low-level feature representations in the auditory cortex via feedback projections (Sussman et al., 2002; Myers & Swan, 2012). This view is further supported by the observations from the crossmodal learning studies, which revealed a task-dependent remapping of the sound representations. It seems that multisensory learning involves multisensory association areas as well as representations of task-relevant non-auditory features, which influence the neuronal responses to sounds through direct or indirect projections to the auditory cortex. Future studies are needed to explore the functional interactions between different processing levels during learning, preferably by means of effective connectivity analyses.

The findings and implications of this thesis should always be considered within the scope of the employed acoustic features and training paradigm. Training was always based on pitch categories and whether other

perceptual categories (e.g. based on velocity) would lead to comparable results is subject to future research. Furthermore, it is noteworthy that low-level feature abstractions in the auditory cortex serving categorization may be limited to perceptual categories with intrinsic physical correspondence. Highly abstract categories (e.g. tool sounds), consisting of arbitrary and highly variable sounds, which are linked only through a semantic or functional association such as the sounds of hammer, saw, and drill may require higher-order (action-related) processes in PFC (Lewis et al., 2005).

The studies presented in this thesis investigated fundamental processing operations in the human auditory cortex and were not particularly designed for clinical applications. Nevertheless, it should be noted that crossmodal plasticity plays a major role during sensory deprivation. Knowledge about the time course and differences with respect to the processing modalities could be considered relevant with respect to the development and implementation of neuroprosthetics. Predicting the success of cochlea implants (CI) for example highly depends on the level of reorganization taking place after sensory loss (Bavelier and Neville, 2002). Input from the visual domain for instance could impair auditory functioning before a CI could effectively prevent cortical remapping (Lee et al., 2001).

The close correspondence of spectro-temporal processing in the auditory and tactile modality and their early convergence in the auditory belt may on the other hand bring the standard of tactile-auditory substitution techniques to a new level. Sensory substitution relies on the transfer of information across modalities, eventually leading to 'perception' in absence of sensory input in the respective modality (for review, see Bach-y-Rita & Kercel, 2003). While crossmodal transfer of information can be achieved through association at some higher-order abstract level (Kim & Zatorre, 2010), 'auditory percepts' of tactile information could potentially be conveyed via early levels at the feature-processing stage. The crossmodal remapping after short-term audio-tactile association training builds the basis for a multimodal representation, which could potentially be exploited by either modality after learning. Whether the learning-induced abstraction of discrete categories from continuous physical input would transfer from the

auditory to the somatosensory domain, eventually leading to discrete percepts of tactile categories after auditory learning needs to be tested.

REFERENCES

Bach-y-Rita P, W. Kercel S (2003) Sensory substitution and the human–machine interface. Trends Cogn Sci 7: 541–546.

Bavelier D, Neville HJ (2002) Cross-modal plasticity: where and how? Nat Rev Neurosci 3: 443–452.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403: 309–312.

Boettiger CA, D'Esposito M (2005) Frontal networks for learning and executing arbitrary stimulus-response associations. J Neurosci 25: 2723–2732.

Brosch M, Selezneva E, Scheich H (2005) Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J Neurosci 25: 6797–6806.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. Science 276: 593–596.

David S V., Fritz JB, Shamma S a (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc Natl Acad Sci U S A 109: 2144–2149.

Desai R, Liebenthal E, Waldron E, Binder JR (2008) Left posterior temporal regions are sensitive to auditory categorization. J Cogn Neurosci 20: 1174–1188.

Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, Ritter W, Murray MM (2002) Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J Neurophysiol 88: 540–543.

Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci 6: 1216–1223.

Fu K-MG, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE (2003) Auditory cortical neurons respond to somatosensory stimulation. J Neurosci 23: 7510–7515.

Goudbeek M, Swingley D, Smits R (2009) Supervised and unsupervised learning of multidimensional acoustic categories. J Exp Psychol Hum Percept Perform 35: 1913–1933.

Guenther FH, Husain FT, Cohen MA, Shinn-Cunningham BG (1999) Effects of categorization and discrimination training on auditory perceptual space. J Acoust Soc Am 106: 2900–2912.

Hackett TA, Smiley JF, Ulbert I, Karmos G, Lakatos P, De la Mothe LA, Schroeder CE (2007) Sources of somatosensory input to the caudal belt areas of auditory cortex. Perception 36: 1419–1430.

Iguchi Y, Hoshi Y, Nemoto M, Taira M, Hashimoto I (2007) Co-activation of the secondary somatosensory and auditory cortices facilitates frequency discrimination of vibrotactile stimuli. Neuroscience 148: 461–472.

Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. Neuron 48: 373–384.

Kim J-K, Zatorre RJ (2010) Can you hear shapes you touch? Exp Brain Res 202: 747–754.

Kral A, Hartmann R, Tillein J, Heid S, Klinke R (2002) Delayed maturation and sensitive periods in the auditory cortex. Audiology & neuro-otology 6: 346–362.

Von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud A-L (2005) Interaction of face and voice areas during speaker recognition. J Cogn Neurosci 17: 367–376.

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255: 606–608.

Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci 30: 7604–7612.

Lee DS, Lee JS, Oh SH, Kim S-K, Kim H-W, Chung J-K, Lee MC, Kim CS (2001) Cross-modal plasticity and cochlear implants. Nature 409: 149–150.

Leech R, Holt LL, Devlin JT, Dick F (2009) Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. J Neurosci 29: 5234–5239.

Levänen S, Jousmäki V, Hari R (1998) Vibration-induced auditory-cortex activation in a congenitally deaf adult. Curr Biol 8: 869–872.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. J Neurosci 25: 5148–5158.

Li S, Mayhew SD, Kourtzi Z (2009) Learning shapes the representation of behavioral choice in the human brain. Neuron 62: 441–452.

Little DM, Thulborn KR (2005) Correlations of cortical activation and behavior during the application of newly learned categories. Cogn Brain Res 25: 33–47.

Myers EB, Swan K (2012) Effects of category learning on neural sensitivity to non-native phonetic categories. J Cogn Neurosci 24: 1695–1708.

Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14: 474–480.

Nordmark PF, Pruszynski JA, Johansson RS (2012) BOLD responses to tactile stimuli in visual and auditory cortex depend on the frequency fontent of stimulation. J Cogn Neurosci 24: 2120–2134.

Ohl FW, Scheich H, Freeman WJ (2001) Change in pattern of ongoing cortical activity with auditory category learning. Nature 412: 733–736.

Pascual-Leone A, Amedi A, Fregni F, Merabet LB (2005) The plastic human brain cortex. Annu Rev Neurosci 28: 377–401.

Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, Sams M (2005) Primary auditory cortex activation by visual speech: an fMRI study at 3T. Neuroreport 16: 125–128.

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12: 718–724.

Ro T, Ellmore TM, Beauchamp MS (2012) A neural link between feeling and hearing. Cereb Cortex 23: 1–7.

Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cogn Brain Res 14: 187–198.

Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC (2001) Somatosensory input to auditory association cortex in the macaque monkey. J Neurophysiol 85: 1322–1327.

Schürmann M, Caetano G, Hlushchuk Y, Jousmäki V, Hari R (2006) Touch activates human auditory cortex. NeuroImage 30: 1325–1331.

Sussman E, Winkler I, Huotilainen M, Ritter W, Näätänen R (2002) Top-down effects can modify the initially stimulus-driven auditory organization. Cogn Brain Res 13: 393–405.

Weinberger NM (1995) Dynamic regulation of receptive fields and maps in the adult sensory cortex. Annu Rev Neurosci 18: 129–158.

SUMMARY

The human auditory system is ideally suited to process and differentiate between the vast acoustic diversity of natural sounds. Importantly though, not all acoustic differences are actually meaningful. Some physical variations in the sound wave might solely reflect distortions of the original signal and therefore do not contribute to the identification of the sound. Moreover, depending on previous experience and the current context, certain spectro-temporal sound features might be more relevant than others to recognize the sound and respond to it appropriately. Processing sounds in an efficient and goal-directed manner therefore requires an abstraction mechanism, which transforms the neural representation of the physical sound characteristics into a behaviorally meaningful perceptual representation, which is robust to irrelevant fluctuations in the signal. In response to rapid changes in situation and task, these perceptual representations must remain plastic to accommodate new environmental demands.

This thesis aimed at achieving a better understanding of the short-term plasticity of neural sound representations in the context of perceptual abstraction of acoustic information. On that account, the research described in chapter 2, 3 and 4 employed controlled learning paradigms to investigate experience-related transformations of neural sound representations. By means of fMRI, neural sound representations could be compared prior to and after successful learning to form two separate categories of artificial ripple sounds. Grouping many different sounds into the same category requires extracting the relevant and suppressing the irrelevant sound features and thereby promotes the formation of a more abstract sound representation. In natural environments, sound categories are seldom learned by means of explicit feedback but rather through the co-occurrence of information in different modalities. Therefore, the empirical studies presented in chapter 3 and 4 incorporated visual as well as tactile stimulation during the learning procedure to investigate the influence of multimodal experience on the neural representation of sound categories.

Chapter 1 commences with an introduction into the concept of categorical perception – the discrepancy between physical and perceptual similarity - and briefly reviews existing literature on neural representations of sound categories. As particular focus of this thesis, the role of learning and specifically multisensory learning in short-term plasticity is discussed.

Following this introduction, **chapter 2** contains the report of the first experimental study. This study assessed whether short-term learning would lead to the formation of discriminable representations of novel sound categories at the level of the auditory cortex. The results revealed that abstract pitch categories could successfully be decoded from distributed activation patterns in the auditory cortex after two days of category training, while the same sounds elicited indistinguishable responses before. This finding provides evidence for the existence of a low-level abstraction process in early human auditory cortex.

Having demonstrated that neural sound representations are amenable to past experiences even beyond the scope of the task, **chapter 3** employed a multisensory learning paradigm to investigate the lasting effect of relevant visual and tactile information on subsequent sound processing. The imaging data revealed a learning-induced crossmodal reorganization of the sound representations. Audio-visual association training led to pronounced sound activations in higher-order visual areas, which did not respond to the same sounds prior to learning. Audio-tactile training on the other hand increased sound processing in the vicinity of the auditory cortex including secondary somatosensory and auditory-somatosensory integration areas. The multivariate analysis of the sound activation patterns described in **chapter 4** additionally revealed that the sound responses in these auditory and tactile regions were category selective. Together, these findings indicate that visual and tactile experiences differentially affect how sounds are processed later in isolation. The close feature correspondence between auditory and vibrotactile stimuli might facilitate direct low-level integration of multimodal information and thereby support the task-dependent abstraction of acoustic information.

The empirical work presented in this thesis provides a marked demonstration for the short-term plasticity of neural sound representations in humans in response to changes in environmental demands. The studies revealed that the *where* and *how* of sound processing do not exclusively depend on the acoustic makeup of the signal but rather on the dynamic integration of current and previous sensory experience as well as prior knowledge.

ACKNOWLEDGEMENTS

Feeling gratitude and not expressing it
is like wrapping a present and not giving it.

~ William Ward

For years, I have read innumerable pages of gratitude, filled by generations of CN PhD students, addressed to a gigantic network of people, inside and outside of University. Honest words, sentimental expressions, deepest sympathy, sincere and blunt from the bottom of their hearts. They praised their supervisors, thanked their advisors and collaborators and worshiped true friendship. And… so will I! Because this thesis is no solo-accomplishment and I am more than happy about the variety of fingerprints left by so many people in the process, bearing witness of the many minds that contributed to this piece of work. From professional advice to emotional encouragement, without the persistent support from others, this thesis would not exist and now is the time and space to thank you for that! And for those, who know me well… I will also be no exception when it comes to sentiments. ☺

Dear Elia, thank you for your trust in my ideas, my abilities and my work in general! When I struggled to believe in myself, you did not give rise to a single doubt. Even in the darkest moments, when all technical equipment seemed to fail on me, you learned to interpret my words in a positive manner: "When you say your data looks 'not too bad', it must be GREAT!!!". I appreciate your enthusiasm and your honest attitude (also when it comes to the weirdest Swiss dessert (MERINGUE with CHEESE!?)! You spared no effort or expenses and always found a way such that I could do my research and I am extremely grateful for these opportunities!

Dear Jean, thank you so much for giving me the opportunity to pursue my own research ideas. Thanks to your open mind, we could hurl ourselves in all kinds of scientific adventures, combining psychophysics with fMRI and MVPA, learning with auditory, visual and tactile stimuli. At the same time I could count on your critical analysis of my behavioral designs and gain from your professional advice.

I owe a lot of gratitude to my third 'ghost-supervisor', Peter! Peter, talking to you about my research, about teaching, about life in general often made me see things clearer. I am so thankful for your patience and all the time you spent listening to my complains and doubts but also all the time we

spent preparing the 'Perception & Attention' group meetings, enthusiastically discussing the big discoveries of 1984 (I will miss that!). Your dedication to science and your endurance inspired me from my first days in Maastricht during the Research Master until now and I hope to experience more of your international dinner parties in the future.

Many thanks also go to Lars Riecke, who's rigorous but always so congenial working attitude has had a large influence on my scientific success. You taught me to spare no effort to construct the right experimental design, as it will save you lots of trouble later on. I am grateful for what I learned under your supervision during the Master thesis and also from your inexhaustible thirst for knowledge and detailed understanding during our Auditory Group meetings.

I want to thank my collaborators, Giancarlo and Amanda, who opened up new exciting possibilities for my research. Giancarlo, you readily shared your wisdom on MVPA and became in turn often the target of my complains about bugs in the code, immature analysis strategies and statistical nonsense. I hope you know how much I appreciate your ever so friendly attitude and how much I admire your knowledge. Your input made this research much more exciting and I am very grateful for that. Amanda, thanks to your expertise, my experiments gained a third sense. We spent weeks (and week-ends) struggling with the logic of the piezostimulator software, time measurements and hardware set-ups, but in the end, we succeeded and due to your help, we are 'picking up good vibrations'. ☺

In this respect, many thanks also go to the Instrumentation team, especially René Finger, Erik Bongaerts, and Johan Gielissen. Thank you for your technical support, which was needed very much! Thanks for not hiding away, when I stood at your door every other day with another complaint about dysfunctional equipment or request for exceptional hardware constructions. My thank also goes to Lotty Huijboom, who often needed to bear my frustrations caused by scanner downtimes, coil dysfunctions, data artifacts, or hardware failure. You always tried your best to accommodate my busy scanning schedule and the required maintenance and made me suffer a little bit less. ☺ Thanks also to the scannexus team (especially

Margo and Esther), who did everything in their power to support me in my impossible attempt to scan 8 subjects (5 sessions each!) in 2 weeks, struggling with very experimental hardware settings.

Lars, although we followed the same road (starting as Coxis in Osnabrück to CN Research Master students in Maastricht), our paths never really crossed until about 5 years ago. Since then, our friendship continuously grew and reached a level, where I can be sure that even if we choose different routes in the future, we will not loose track of each other again. Your friendship means so much to me and I am grateful for all the wonderful time we spent (and will spend) together, playing board games, drinking wine (or tea!), climbing, and on top of everything, TALKING! Thank you for sharing so many emotional moments with me, from excessively enthusiastic over maniac to deep depression. You know that all this is part of who I am and you live these moments with me! Apart from being a great friend, you are a terrific colleague, too and I am very thankful for your essential contribution to the studies in this thesis (beyond MVPlab!).

Roberta, you went through many adventures since you came to Maastricht: swinging from tree to tree, biking in the rain, eating garlic (even on PIZZA!!!) and all that while being surrounded by Germans! Thank you so much for your incredible support in so many nerve-wracking situations!!! I will miss your straight and genial attitude very much and I hope that my (future) kids will get the chance to pick up some southern and northern Italian style from you and Giulio in Lausanne (or wherever you will be) as you're the most natural person around kids I've ever met!

Felix, what can I say… I probably would have worked (A LOT) more without you being around… BUT then I would have missed not only our in-depth discussions about stimulus designs and your incredible presentation skills, but more importantly, our (sometimes hours-long) discussions about politics, society, family, friendship, parentship, gardening, cooking, furniture, even sports (although my contribution definitely lacked competence when I mediated between you and Sven after Bayern München or VfB games)... and who knows what else. We counseled each other in all kinds of vital matters and your opinion was always very important to me!

Thank you soooo much not only for listening but for understanding and most importantly for your honest assessment and advice! Thank you for all the wonderful cooking nights (which I insist on keeping up no matter how far the distance between our houses) and the wonderful balance between challenge and laisser-faire!

Thank you, Gesa for your uncomplicated and friendly nature, for being the perfect girl for Felix, for being absolutely hilarious when you're drunk, for teaching us the right environmental discipline (i.e. collecting lemons on the beach in the middle of the night), for teaching me how to lead horses with my posture only, for exchanging all the wonderful childhood memories of small town girls.

Dear Britta, sitting not more than 1m apart (in two adjacent offices), we often shared our frustration via skype ("… zum 6. Mal…", "7!!!!"). Thanks for being so compassionate, for great climbing evenings with me as newbie, for 'TODES-leckere' Abendessen and 'MEGA-nette' Gespräche and the introduction into Öcher Platt! I hope to visit you soon in Cape Town (although that also means not having you around any longer).

I owe special thanks to my oldest Maastricht friends, Michelle and Tom. Although there's no way I could ever reach your level of energy, diligence and endurance (Michelle) or your persuasiveness, linguistic talent, and spontaneity (Tom), you two always inspired and stimulated me and your company let me aim high and believe in myself! Michelle, thank you for your lasting friendship, for sticking with me all this time (even when the Vampire fever hit me), balancing my temper! Tom, there will never be a time, when you don't drive me CRRRRAAAZY, but I guess that's the way it should be.

Dear Nick and Martin, you were the first people I talked to at UM. Intimidated by the modern and professional appearance, when I came from Osnabrück for the interview, you two convinced me that the Research Master was THE way to go. Your friendly attitude and the enthusiastic and extrovert performance of Rainer in the interview had an immense impact on my decision. Thank you Martin for eagerly sharing your knowledge on localizers, book printing, Australia, or any kind of fact you picked up yet

again. Nick, long before your insightful “Freundezusammenführung”, I appreciated your company and your down-to-earth assessment of attitude and ambition of CN Research Master students. Thank you for always listening to my doubts and sharing yours. See you soon…. In COLOGNE! ☺

Thanks to the whole Auditory Group (whose members are with some fluctuations): Elia, Milene, Giancarlo, Federico, Lars R., Lars H., Fren, Noel, Julien, Michelle, Nick, Roberta, Sanne, and Kiki. I very much enjoyed our (not so) regular meetings. They were always full of laughter and so nicely fuss-free, such that they provided a platform for outright discussions about innovative ideas, ambitious designs and puzzling results. Thanks also to the newly established Multisensory Group (i.e. Nienke, Judith, Nina, Gojko, Sanne, Katie, Kiki, and Amanda), who supported me during the final phase of my PhD. It is really too bad, that I could not join more of those inspiring meetings.

Dear Judith, Joost, and Armin, thank you for your reliable support with my favorite software. Learning studies come with large amounts of data and with your help, I conquered the world of scripting which often saved me days of work.

Thank you, Annemie, Christl, and Riny! Your on-site support was very valuable to me and I appreciated your uncomplicated and friendly attitude. Dear Saskia Diesbergen, although we only met once in person, I thank you so much for your absolutely essential support from Tilburg! You were my #1 contact person in Tilburg and you never left any of my questions unanswered or any of my administrative matters unsolved.

Dear Jeannette, you basically welcomed me as employee at Maastricht University in Leo’s group. Thank you for the many hours, we spent together in the Allegra lab with brave and not-so-brave children, lots of tea and good chats. Thanks for your lasting friendship and your always so positive attitude. Dries, Anniek, and Gonny, thank you, too for your warm welcome in the dyslexia group, for sharing your wisdom (scientific and life-wise) with me!

Thank you, Valentin for the many many rides to and from Aachen and Maastricht. You are probably the most civilized person I have ever met!

Next to you, I sometimes felt like the little girl from the gutter. Thanks for politely listening to my swearing and never complaining about my driving style! ☺

Thank you, Inge and Thomas for being such wonderful and tolerant roomies. After saving you from the squeaky thermos bottle (sorry Mario), you patiently endured my ups and downs!

Thanks to all the nice colleagues and friends who spent their valuable working time in the UNS40 basement, supporting me during my numerous scanning sessions! Your presence was not only indispensable for the success of my studies, it also saved me some very lonely and boring hours and I am glad to have spent them with you instead.

Thank you Bernadette for being a great mentor, for your ever so straightforward and kind nature and for teaching me that failure is part of life.

Finally, I honestly want to thank all the great people who create such an exceptional collegial atmosphere in the CN department. As much as I am looking forward to the future with all its changes it is going to bring, I will bitterly miss the positive casual attitude with its mutual support, the friendly advice and constructive criticism, the affinity for good coffee (Bandito!), good food (yes, I DO like bitterballen), good parties and not so good party locations (THE ALLA), and all the good times we had in Maastricht, Barcelona, Quebec City, Lausanne, … thank you Federico (for not stealing Sven away from me. ☺ I will miss your parties and other great evenings at your place.), Anna G. and Matteo, (for the great cooking events!), Marin and Bert (I hope that we'll find the time for more bouldering in the future!), Job, Francesco, Marieke, Aline, Betty, Rosanne, Joao, Martin H., Jan, Sven, Sanae, Christianne, Franzi, Kamil, Teresa, Eva, Mario, Arne.

Julia, thank you for holding onto our friendship also in difficult times! I am very happy that we will share many more experiences life brings in our future.

Dear Becca, you are the true "Öcher Mädche" and for my sympathy for Aachen and you, I am extremely happy about your future plans of turning your back on Kiel and eventually returning to your birthplace. Thank you for

your outspoken nature, for the wonderful evenings in 'our' winebar, Spargelessen in the Brasserie Aix, for your valuable friendship!

Dear Svenja and Uta, I am very thankful for your friendship through all these years despite the distance between our homes. Let's keep up the rotation! ☺

Thanks to all the great friends who accompanied me through this very important period of my life, thank you for listening to my stories, for your honest interest in my research, for trying to understand that this is a *real* job, for distracting me when it was needed, thanks for nice weekends in Aachen, Stuttgart, Zürich, Düsseldorf, Hamburg, and Nürnberg, thanks to Marcus, Johannes, Claudi, Jan, Kaddie, Steffen, Franzi, Oli, Verena, Reini, Nadja, Saskia, Karsten, Caro, Matthias and Stephan.

Liebe Karin, lieber Frank, dies ist nun ein weiterer großer Abschnitt meines Lebens, den ihr von Anfang bis Ende begleitet habt. Wer hätte das gedacht vor 12 1/2 Jahren, dass ihr die Schülerin aus Braunschweig als Schwiegertochter bis zur Promotion begleitet. ☺ Danke, dass ihr mich all die Zeit so toll unterstützt und - auch wenn es dem ehemaligen Betriebsrat bestimmt gegen den Strich geht - Verständnis für meine Arbeitseinheiten an Wochenenden und im Urlaub aufbringt. Euch hab ich zu verdanken, dass sich meine Kollegen über mein "kansch", "weisch", "siehsch" lustig machen, und ich bin stolz auf meine schwäbische Familie! Danke für die Einführung in richtig gute Musik, in die schwäbische Küche und für lustige Kniffelabende!

I owe special thanks to my parents, who not only kindly listened to my stories about Cognitive Neuroscience but developed a passion for it on their own! Your genuine interest and empathy is fantastic and absolutely exceptional and I am very thankful for that. Thank you, Papa, for sharing your own experience and wisdom, for hours of telephone counseling when I was at the edge of despair but also for you compassionate joy when I was brimming over with mirth due to some small accomplishment. Thank you for truly trying to understand my paper and thank you for giving up after days of re-reading it. ☺ I inherited your ambition, your curiosity and your eye for detail, which helped me very much to accomplish this thesis. Thank

you for challenging me (almost) every day with the newest press releases in the field of Neuroscience and thank you for tolerating my (often) withering assessments. Thank you, Mama, for your care, for making me feel at home whenever I visit, for your patience and your openness! Thank you both so much for your unconditional support but also for openly sharing your honest opinions and judgments, for helping me to make my own decisions. Finally, I want to thank my big brothers, Jan and Dirk, for completing this wonderful family that gives me absolutely essential support on this journey through life.

Sven, writing this means moving on, leaving behind 5 wonderful years, we spent together in Aachen. Although this makes me very sad, this is (only) about 40% of the time we already know each other and only about 6% of the time I am planning on spending with you. ☺ I am thrilled to meet the new challenges and surprises waiting for us, to find *our* new places, *our* new routines, *our* new moments and cherish the ones that make us 'Sven und Anke'. Thank you for reminding me of who I am and allowing me to change at the same time! Your indispensable support during this tough time is not measurable in deeds; it is your part in *us* that gave me the strength and faith to complete this journey. You are incredible, you are my everything, I love you so much!

CURRICULUM VITAE

Anke Ley (née Walter) was born in Los Alamos, New Mexico, USA on 9 December 1983. Growing up in Germany, she graduated with the secondary school certificate in 2003 at the Gymnasium am Löwenwall in Braunschweig. Anke pursued her education studying Cognitive Science at the University of Osnabrück and spent the winter semester 2005 at the State University of New York at Oswego, USA. She was awarded with the degree of Bachelor of Science (cum laude) in 2006. Her bachelor thesis on overt visual attention is based on empirical research conducted under the supervision of Prof. Peter König and Dr. Selim Onat (University Osnabrück). From 2006 till 2008, Anke followed the Research Master program in Cognitive Neuroscience at Maastricht University. Within the scope of her master thesis supervised by Prof. Dr. Elia Formisano and Dr. Lars Riecke, Anke investigated the neural basis of the auditory continuity illusion. Anke graduated (cum laude) in 2008 and continued working at the Department of Cognitive Neuroscience in the group of Prof. Dr. Leo Blomert, investigating dyslexia in children using EEG and fMRI. In 2009, she started the PhD program at Tilburg University in close collaboration with the Department of Cognitive Neuroscience, Maastricht University. From October 2013, Anke will work as Referee at the Deutsche Forschungsgemeinschaft.

PUBLICATIONS

Articles

Bonte, M., Frost, M., Rutten, S., **Ley, A.**, Formisano, E., & Goebel, R. (2013). Development from childhood to adulthood increases morphological and functional inter-individual variability in right superior temporal cortex. NeuroImage 83: 739-750.

Ley, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., & Formisano, E. (2012). Learning of new sound categories shapes neuronal response patterns in human auditory cortex, J Neurosci 32: 13273-13280.

Riecke, L., **Walter, A.**, Sorger, B., & Formisano, E. (2011). Tracking vocal pitch through noise: Neural correlates in nonprimary auditory cortex. J Neurosci 31: 1479-1488.

Ley, A., Kaas, A., Formisano, E., Vroomen, J. (in preparation). Audio-tactile and audio-visual learning promote crossmodal reorganization at different neural processing levels.

Ley, A., Hausfeld, L., Kaas, A., Vroomen, J., Formisano, E. (in preparation). Learning of new sound categories: distinct neural effects of visual and tactile contextual information.

Conference publications

Bonte, M., Frost, M., Rutten, S., **Ley, A.**, Formisano, E., & Goebel, R. (2012). *Inter-subject variability in right superior temporal cortex increases with development from childhood to adulthood*. Poster session presented at 4th International Conference on Auditory Cortex, Aug 31-Sep 3, Lausanne, Switzerland.

Ley, A., Hausfeld, L., Vroomen, J., Valente, G., De Weerd, P., & Formisano, E. (2011). *Distributed cortical representations of newly-learned auditory categories*. Oral presentation at the 17th Meeting of The European Society for Cognitive Psychology (ESCOP), Sep 29-Oct 2, San Sébastian, Spain.

Walter, A., Vroomen, J., Hausfeld, L., Valente, G., De Weerd, P., & Formisano, E. (2011). *Distributed cortical representations of newly-learned sound categories*. Poster session presented at 17th Annual Meeting of the Organization on Human Brain Mapping (HBM), Jun 26-30, Québec City, Canada.

Riecke, L., **Walter, A.**, Sorger, B., & Formisano, E. (2009). N*eural representations of illusory voice continuity in auditory cortex*. Poster session presented at 3rd

International Conference on Auditory Cortex, Aug 29-Sep 2, Magdeburg, Germany.

Froyen, D., Willems, G., **Walter, A.**, & Blomert, L. (2009). Cross-modal effects of letters on the MMN to speech sounds reveals deficient letter - speech sound processing in children with dyslexia. Front. Hum. Neurosci. Conference Abstract: MMN 09 Fifth Conference on Mismatch Negativity (MMN) and its Clinical and Scientific Applications.

Walter, A., Riecke, L., & Formisano, E. (2008). *Voice continuity illusion: Neural mechanisms of filling-in complex auditory objects*. Oral presentation at the 7th Endo-Neuro-Psycho (ENP) Meeting, Jun 4-6, Doorwerth, The Netherlands.

Walter, A., Riecke L., & Formisano, E. (2008). *Voice continuity illusion: Neural mechanisms of filling-in complex auditory objects*. Poster session presented at the 1st Cognitive Neuroscience Conference Nijmegen (CNCN), May 13-14, Nijmegen, The Netherlands.